Master's thesis
Experimental particle physics

b jet response studies at the CMS experiment at LHC

Antti Pirttikoski

August 16, 2021

Supervisor(s): Associate Professor Mikko Voutilainen

Censor(s): Associate Professor Mikko Voutilainen
Minsuk Kim

UNIVERSITY OF HELSINKI
TCM Master's Programme

P.O. Box 62 (Gustaf Hällströmin katu 2)
FI-00014 University of Helsinki

HELSINGIN YLIOPISTO — HELSINGFORS UNIVERSITET — UNIVERSITY OF HELSINKI

Tiedekunta — Fakultet — Faculty	Koulutusohjelma — Utbildningsprogram — Education programme
Faculty of Science	TCM Master's Programme

Tekijä — Författare — Author
Antti Pirttikoski

Työn nimi — Arbetets titel — Title

b jet response studies at the CMS experiment at LHC

Opintosuunta — Studieinriktning — Study track
Experimental particle physics

Työn laji — Arbetets art — Level	Aika — Datum — Month and year	Sivumäärä — Sidoantal — Number of pages
Master's thesis	August 16, 2021	86 pages

Tiivistelmä — Referat — Abstract

LHC is the highest energy particle collider ever built and it is employed to study elementary particles by colliding protons together. One intriguing study subject at LHC is the stability of the electroweak vacuum in our universe. The current prediction suggests that the vacuum is in the metastable state. The stability of the vacuum is dependent on the mass of the top quark, and it is possible that more precise measurement of the mass could shift the prediction to the border of the metastable and stable states.

In order to measure the mass of the top quark more precisely, we need to measure the bottom (b) quarks decaying from it at high precision, as top quark decays predominantly into a W boson and a b quark. Due to the phenomenon called hadronisation, we can not measure the quarks directly, but rather as sprays of collimated particles called jets. The jets originating from b quarks (b jet) can be identified by b-tagging. The precise measurement and calibration of the b jet energy is crucial for top quark mass measurement.

This thesis studies the b jets and their energy calibration at the CMS, which is one of the general purpose detectors along the LHC. Especially the b jet energy scale (bJES) is under the investigation and the various phenomena affecting to it. For example, large fraction of b jets contain neutrinos, which cannot be measured directly. This increases uncertainties related to the energy measurement. Also there are problems how precisely the formation and evolution of the b jets can be modelled by Monte Carlo event generators, such as Pythia8, which was utilized in this thesis.

The aim of this thesis is to evaluate how big effect on the bJES is caused by the various different phenomena, which presumably weaken the precision of the b jet measurements. The studied phenomena are the semileptonic branching ratios of b hadrons, branching ratios of b hadron to c hadron decays, b hadron production fraction and parameterization of the b quark fragmentation function.

The combined effect of all four different rescaling features mentioned above, suggests that bJES is known at 0.2% level. A small shift of -0.1% in the Missing transverse energy Projection Fraction (MPF) response scale is detected at low p_{T} values, which vanishes as the p_{T} increases. This improves remarkably 0.4-0.5% JES accuracy achieved during at CMS during Run 1 of the LHC. However, there are still many ways we can improve the performance presented here. Definitely there is a need for further studies of the rescaling methods before results could be utilized in the corrections of bJES to do precision measurement of the top quark mass.

Avainsanat — Nyckelord — Keywords
CERN, LHC, CMS, jet, energy scale

Säilytyspaikka — Förvaringsställe — Where deposited

Muita tietoja — övriga uppgifter — Additional information

HELSINGIN YLIOPISTO — HELSINGFORS UNIVERSITET — UNIVERSITY OF HELSINKI

Tiedekunta — Fakultet — Faculty	Koulutusohjelma — Utbildningsprogram — Education programme
Matemaattis-luonnontieteellinen tiedekunta	TCM-maisteriohjelma

Tekijä — Författare — Author
Antti Pirttikoski

Työn nimi — Arbetets titel — Title
b-jettien responssitutkimus CMS-kokeella

Opintosuunta — Studieinriktning — Study track
Experimental particle physics

Työn laji — Arbetets art — Level	Aika — Datum — Month and year	Sivumäärä — Sidoantal — Number of pages
Pro gradu	12. elokuuta 2021	86 sivua

Tiivistelmä — Referat — Abstract

LHC hiukkaskiihdytin on korkeaenergisin hiukkaskiihdytin koko maailmassa ja sitä käytetään alkeishiukkasten tutkimiseen törmäyttämällä protonisuihkuja toisiinsa. Yksi mielenkiintoisimmista tutkimusaiheista LHC:llä liittyy sähköheikon tyhjiön stabiiliuteen. Nykymittausten valossa tyhjiön ennustetaan olevan metastabiilissa tilassa. Tämä ennuste perustuu laskelmaan, joka riippuu huippu-kvarkin (t-kvarkki) massasta. Tarkempi t-kvarkin massan määritys voisi siirtää ennustetta lähemmäksi metastabiilin ja stabiilin tyhjiön raja-aluetta.

Huippukvarkin massan määrittäminen on suoraan riippuvainen pohja-kvarkin (b-kvarkki) mittauksen tarkkuudesta, koska t-kvarkki hajoaa melkein pelkästään W-bosoniksi ja b-kvarkiksi. Hadronisaatioksi kutsutun ilmiön seurauksena emme pysty havaitsemaan kvarkkeja suoraan. Sen sijaan havaitsemme niiden synnyttämiä hiukkasryöppyjä, joita kutsutaan jeteiksi. B-kvarkista syntynyttä jettiä kutsutaan puolestaan b-jetiksi ja näiden tarkka mittaaminen on erityisen tärkeää t-kvarkin massan määrityksessä.

Tässä työssä tutkimme b-jettien energiaskaalaa CMS-kokeessa, joka on yksi LHC:n yleisluontoisista kokeista. B-jetin energiaskaalaan vaikuttaa esimerkiksi neutriinoiden suuri osuus b-jeteissä sekä simulaatio-ohjelmien rajallinen tarkkuus, joilla protoni-protoni törmäyksiä mallinnetaan. Tutkimme erityisesti neljää eri mekanismia, jotka vaikuttavat b-jettien energiaskaalaan. Nämä mekanismit ovat b-hadronien semileptoninen hajoaminen, b-hadronien hajoaminen c-hadroneiksi, b-hadronien tuotanto-osuudet ja b-kvarkkien fragmentaatiofunktiot.

Näiden neljän mekanismin yhteisvaikutus b-jettien energiaskaalan tarkkuuteen on noin 0.2%. Myös negatiivinen 0.1% suuruinen muutos havaitaan MPF-responssissa (Missing transverse energy Projection Fraction response) matalalla energia-alueella. Tämän työn tulokset parantavat huomattavasti 0.4-0.5% tarkkuutta, joka on aikaisemmin saavutettu b-jetin energiaskaalan mittauksissa CMS:llä LHC:n ensimmäisen käyttövaiheen aikana. Tämän työn käytetyissä metodeissa on kuitenkin vielä paljon jatkokehitystarvetta ja aihetta lisätutkimuksille ennen kuin esiteltyjä tuloksia voitaisiin käyttää hyödyksi t-kvarkin massan mittauksessa.

Avainsanat — Nyckelord — Keywords
CERN, LHC, CMS, jetti, energiaskaala

Säilytyspaikka — Förvaringsställe — Where deposited

Muita tietoja — övriga uppgifter — Additional information

Contents

1. Introduction

The Standard Model (SM) describes the most fundamental elementary particles and their interactions. In order to study these fundamental particles, we need extremely high energies. One way to achieve high enough energies, is to collide particles together nearly at the speed of light. This is exactly how the Large Hadron Collider (LHC) at CERN works, as it accelerates and collides protons together. The greatest achievement of the LHC so far, has been the detection of the Higgs boson in 2012, which was the last missing piece to SM [1, 2].

One of the four main detectors at the LHC is the Compact Muon Solenoid (CMS). The CMS is a highly complex detector system consisting of various sub-detectors laid on top of each other throughout its 15 m wide diameter. The CMS can detect multiple types of particles, one of which is called hadrons. The hadrons are made from quarks, which in turn cannot be detected directly as individual particles. This is known as color confinement, and it is caused by the strong interaction of the SM. When the quarks are produced in the high energy collisions, a phenomenon called hadronisation takes place, in which quarks and gluons produce collimated sprays of particles known as jets.

Thus in order to study the quarks, we need to first reconstruct the jets. We are interested in the four-momentum of jets, as it is directly related to the four-momentum of the quark or gluon is was originated from. As the jet angles are generally well measured, the largest uncertainty comes from jet energy or magnitude of the transverse momentum p_{T}. In general the measured and reconstructed jet energy do not match with the real energy of the jet at the particle level. This is due to the limited resolution of the detectors and the limitations of the event reconstruction. For these reasons, the jets need to be calibrated. The jet energy calibration aims to shift the jet energy to the correct jet energy scale (JES) by applying series of jet energy corrections (JEC) [3].

One of the most interesting aspects of the precise jet calibration is related to the measurement of the top quark mass (m_t). The top quark will decay almost exclusively to a bottom quark (b) and a W boson. Thus the calibration of the jets originated from the b quarks, called as b jets, is essential in the determination of the m_t. The special interest in the m_t is due to the fact that it predicts with the mass of the Higgs boson (m_H) the state of electroweak (EW) vacuum of our universe. The current measurements suggest that the EW vacuum minimum is in the metastable state. However, the precision in which the m_t is known restricts accuracy of this prediction, and it is possible that vacuum minimum could shift almost to the border of metastable-stable states within the current uncertainty

bounds [4].

This thesis studies the calibration of the b jets and various phenomena which affect it. More specifically, the thesis tries to evaluate how big effect these phenomena have on the precision of the JES of the b jets (bJES). This is done by using Monte Carlo (MC) event generator called Pythia 8 (P8) [5]. The MC generators are used in the particles physics to simulate the events that take place in the colliders.

The phenomena studied in this thesis are the semileptonic branching ratio of b hadrons, b-to-c hadron branching ratios, b hadron production fraction and parameterization of the b quark fragmentation function. These were selected as they presumably weaken the bJES precision. The semileptonic decays mentioned before are decays in which along with the other particles, a neutrino is produced. Neutrinos cannot be directly measured at the CMS, and thus they weaken considerably the bJES precision.

It is also assumed that the P8 cannot model the studied features completely accurately, thus the b jet energy differs between the P8 and experimentally measured data from the CMS. In order to evaluate how big this difference is, two P8 $Z(\rightarrow \mu\mu) + b$ samples are produced. One of them is rescaled so that it corresponds to the experimental data, while another one is not touched. Both samples are then run through a detector simulation, after which the b jet energies are compared to see if the bJES is shifted due to the rescaling.

The thesis is organized in the following way. First, in Chapter 2, the basic conventions used in the high energy collider physics are reviewed. After this, the SM is briefly introduced, followed by inspection of the stability of the EW vacuum. In Chapter 3, the basic steps of the MC event generation are introduced, with emphasis on the hadronisation and the different features that are rescaled in P8. Next the LHC and CMS are presented in Chapter 4, where especially the reconstruction related to jets is handled more thoroughly. Chapter 5 describes the P8 event generation environment and detector simulation used in the thesis, as well as all the rescaling methods. The results of the rescaling are presented Chapter 6 where the shift in the bJES is shown due to the individual rescaling methods and also for the total combination of them. The observed effect are then discussed Chapter 7. The thesis ends with the conclusions in Chapter 8.

2. Theoretical overview

This chapter starts with a short review of the common quantities used in the high energy and collider physics. After this, the main features of the Standard Model are introduced briefly. The chapter ends with a section focusing on the relation of b jets to the top quark mass measurements, and thus indirectly affecting on the measurement of the electroweak vacuum stability. The first section uses Refs. [6, 7, 8], while the section on Standard Model is based on Refs. [8, 9, 10].

2.1 Conventions of high energy collider physics

The system of units used in high energy physics is called *natural units*. The convention is to set the speed of light and the reduced Planck constant to unity, *i.e.* $c = \hbar = 1$, and then express the quantities in terms of energy. The energies are expressed as electronvolts (eV), which is the energy that electron acquires as it is accelerated across 1 Volt potential difference. Thus the energy, momentum and mass can be given in units of eV, and time and length in eV^{-1}.

The electronvolt is far more suitable unit to be used in the high energy physics, as the SI unit Joule is many orders of magnitude smaller, which can be seen from the conversion, 1 eV $\approx 1.6 \times 10^{-19}$ J. The elementary particle mass scales range from eV to GeV, while the highest energy collider operates at the moment in the energy scale of TeV.

If we consider the collision of two elementary particles a and b in the center-of-momentum frame (c.m.) travelling along z-axis, we can express the four-momenta of particles as

$$p_a = (E, 0, 0, E), \quad p_b = (E, 0, 0, -E), \tag{2.1}$$

where we have assumed that masses of the particles are much smaller compared to the energy of the collision and that particles a and b are accelerated with same energy. The total energy available in the c.m. frame is thus $E_{cm} = 2E$. The E_{cm} can also be expressed using a Lorentz invariant variable, called *Mandelstam variable* s, which is defined as

$$s \equiv (p_a + p_b)^2 = E_{cm}^2, \tag{2.2}$$

from which we get $E_{cm} = \sqrt{s}$, which is the usual form used to denote the total collision energy.

Besides energy, the *cross-section* σ is another important variable related to collision experiments. When two particles come close together, there is some probability that they will scatter off or transform to another particles. The cross section depicts this probability of interaction between particles. It can be defined as

$$\sigma = \frac{R}{F}, \tag{2.3}$$

where R is the total rate of scattering events and F the incoming flux of particles, *i.e.* number of particles crossing unit area per unit time. By inspecting the definition of σ, it can be seen that it is given in the units of $[\text{length}]^2$. Thus the σ actually represents the cross-sectional area which the particles have with respect to each other in the collision. Usually the cross-sections of high energy physics are given in the units of *barns* (b), where $1\text{b} = 10^{-28}$ m^2.

In order to calculate how many interactions happens in a given time interval, we need to know the *instantaneous luminosity* $\mathcal{L}$, besides the cross-section of the interaction. For example, if two bunches of particles are collided, the $\mathcal{L}$ is given as [11]

$$\mathcal{L} = \frac{f n_1 n_2}{4\pi \sigma_x^* \sigma_y^*} \mathcal{F}, \tag{2.4}$$

where n_1 and n_2 are number of particles in the colliding bunches, f is the collision frequency, σ_x^* and σ_y^* characterises the root-mean-square horizontal and vertical beam sizes, and $\mathcal{F}$ is a factor of order 1, which takes account different geometric and dynamic effects.

By integrating the $\mathcal{L}$ over the lifetime of the experiment, we get the *integrated luminosity* L_{int}, in other words $L_{\text{int}} = \int \mathcal{L} dt$. Now we can calculate the number of interactions of the given process simply by using the following equation

$$N = \sigma L_{\text{int}}. \tag{2.5}$$

The Equation 2.5 has an important consequence to the collider design. This for the reason that cross-section tends to decrease with increasing energy. Thus, in order to obtain reasonable amount of interactions in high energy colliders, higher luminosities are also required.

The other set of important variables are related to how we measure the position of the particle inside the collider. Usually in the collider experiments, the z-axis points along the counter-clockwise rotating beam direction, x-axis points radially inward to the center of the circular collider, and y-axis vertically upwards. However, it is more common to describe particles using the angles with respect to the beam axis (or z-axis). The azimuthal angle ϕ is measured with respect to x-axis in the x-y plane. The polar angle θ is measured from the z-axis in the z-y plane.

Instead of using the θ-angle, it is more convenient to use a variable that is invariant under the boosts along the z-axis. For example, this is desirable property especially in hadron colliders, as the collision happening in the c.m. frame of the hadrons is not the same c.m. frame in which the constituents

taking part to the actual collision are. A suitable variable for this purpose is called *rapidity* y, which is defined as

$$y = \frac{1}{2}\ln\Big(\frac{E + p_z}{E - p_z}\Big), \tag{2.6}$$

where E and p_z are the energy and z-component of momentum of the particle. However, this form is rarely used, as it can be simplified further by high energy approximation, where the mass is negligible compared to energy, and thus $p_z \approx E\cos\theta$. Putting this into equation 2.6 yields the *pseudorapidity*

$$\eta \equiv -\ln\Big(\tan\frac{\theta}{2}\Big), \tag{2.7}$$

which is only dependent on the θ-angle.

Yet another useful variable is the *transverse momentum* p_T of the particle, defined as $p_\mathrm{T} = \sqrt{p_\mathrm{x} + p_\mathrm{y}}$. As mentioned before, we can assume that the colliding particles have no momentum relative to the x- and y-axes before the collision. Thus, due to the conservation of the momentum, the vector sum over final state particle p_T must be equal to zero, which makes the p_T such an useful variable in the analyses.

Finally, it is sometimes useful to measure the distances between two objects in the ϕ-η plane. A variable commonly used for this purpose is called *radial distance* $\Delta\mathrm{R}$, defined as

$$\Delta\mathrm{R} = \sqrt{\Delta\phi^2 + \Delta\eta^2}, \tag{2.8}$$

which is also invariant under boosts along the z-axis.

2.2 The Standard Model

The *Standard Model* is the most fundamental and precise theory about our universe at the moment. It describes the particles and their interactions, and mathematically it is written in the language of Quantum Field Theory (QFT). More precisely, the SM is a gauge theory based on the symmetry group $SU(3) \otimes SU(2) \otimes U(1)$, which describes strong, weak and electromagnetic interactions, respectively. These interactions can be derived from a single general principle, namely the requirement of local gauge invariance. However, this section describes the SM more qualitatively, and more thorough description of SM in terms of QFT can be found in many standard books, *e.g.* [12].

The elementary particles of SM can be divided into three different categories according to their spin. The spin-1/2 particles are called as *fermions*, or as "matter particles", and they can be further divided into two subcategories, namely *leptons* and *quarks*, both consisting of 6 particles. The spin-1 particles are called *vector bosons*, and they mediate the three different forces mentioned before, *i.e.* strong, weak and electromagnetic interactions. The final elementary particle category, *scalar bosons*, consist of a single spin-0 particle called *Higgs*

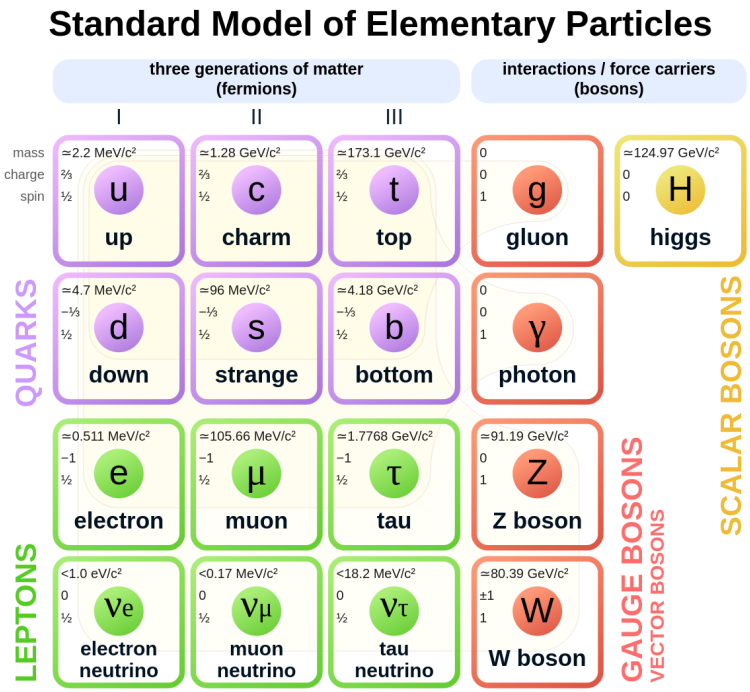

Figure 2.1: All the elementary particles of SM with their corresponding mass, charge and spin [13].

boson (H), which allows the other elementary particles acquire mass. All these elementary particles are shown in Figure 2.1.

As seen in Figure 2.1, the fermions can be put into three generations, each consisting of two quarks and two leptons. The masses of the particles increase between each generation, but otherwise the properties of the particles are the same for each generation. For example, the first generation consists of up (u) and down (d) quarks, which have $Q = +2/3$ and $Q = -1/3$ charges, respectively. In addition, there are electron (e) with $Q = -1$ charge and neutral, very weakly interacting, electron neutrino (ν_e). The same pattern of charges repeats in the second and third generation, where you can find charm (c) and top (t) quark with charges $Q = +2/3$, strange (s) and bottom (b) quarks with $Q = -1/3$, muon (μ) and tau (τ) with $Q = -1$ and corresponding neutral muon (ν_μ) and tau (ν_τ) neutrinos. The fundamental reason for why there exists exactly three generations instead of any other number is unknown.

To complete the picture about the fermions, the several degrees of freedom that each fermion possess have to be taken into account. First of all, each fermion has its own *antiparticle*, which is otherwise identical to the 'normal' particle, but it is oppositely charged, *i.e.* $Q_{\text{antiparticle}} = -1 \cdot Q_{\text{particle}}$. The second degree of freedom is related to the spin of the fermion. Namely, each fermion can also be decomposed into its left- and right-handed pieces. Roughly speaking, the right-handed particles have their spin aligned with its momentum, while left-handed particles have their momentum and spin in opposite directions. The third degree of freedom only concerns quarks. Corresponding to electric charge,

all quarks have something called *color charge*, but unlike electric charge, it has three different values, which are named as red, green and blue.

The vector bosons consist of photons (γ) mediating the electromagnetic interactions, 8 gluons (g) mediating the strong force, and Z and $W^{\pm}$ bosons mediating the weak neutral-current and charged-current interactions, respectively. Of these bosons, gluons and photons are massless, while the Z and $W^{\pm}$ are massive. The theory describing the interactions of photons with other particles is called *quantum electrodynamics* (QED). In order to interact with photons, the particles must have electric charge. Thus all the other fermions expect neutrinos take part to QED processes, and as the photons themselves are neutral, they cannot self-interact.

The theory of gluons interacting with quarks is known as *Quantum Chromodynamics* (QCD). Only particles with color charge can interact with gluons, thus other fermions than quarks do not take part in the QCD processes. An important difference between QCD and QED is that the gluons can self-interact, as they carry the color charge themselves. This has dramatic effects on how the strong force affects to the other particles. It causes a phenomenon called *asymptotic freedom*, where the *strong coupling* α_s approaches zero at high energies, or short distance scales, and correspondingly increases with lower energies and larger distances. This is also known as *running of coupling*, and it leads to *confinement* of quarks and forming of bound states with net-zero color charge, called *hadrons*. The hadrons can be divided into two categories: *mesons* and *baryons*. The mesons consist of quark-antiquark pair, while the baryons are made of three quarks (or three antiquarks).

In addition to massive bosons, the weak force differs from QED and QCD in one important aspect. Namely, it does not respect the symmetry of *parity*. The parity transformation corresponds to spatial inversion through the origin and can be thought as inspecting the phenomena reflected by a mirror. This property of weak force is know as *parity violation* and it ultimately manifest itself by the fact that only the left-handed particles take part in the weak interactions.

The neutral Z bosons interact in the same manner as photons, but they can also couple to neutral particles, like neutrons. The W bosons, on the other hand, are electrically charged, and W^+ and W^- carry $Q = 1$ and $Q = -1$ charges, respectively. Furthermore, the W bosons couple charged leptons to corresponding neutral neutrinos, *e.g.* μ and ν_μ, and all possible quark combinations differing by one unit charge, *e.g.* u and d or $\bar{c}$ and $\bar{b}$.

The probability that two quarks couple via the weak force is given by the 3×3 matrix known as Cabbibo-Kobayashi-Maskava (CKM) matrix. The mixing of d, s and b quarks can be written as matrix equation containing the CKM matrix in the following way

$$\begin{pmatrix} d' \\ s' \\ b' \end{pmatrix} = \begin{pmatrix} V_{ud} & V_{us} & V_{ub} \\ V_{cd} & V_{cs} & V_{cb} \\ V_{td} & V_{ts} & V_{tb} \end{pmatrix} \begin{pmatrix} d \\ s \\ b \end{pmatrix}. \tag{2.9}$$

The absolute values of the CKM matrix elements represent the probabilities

associated with the corresponding quark coupling. The elements are measured experimentally, and are approximately equal to [11]

$$\begin{pmatrix} |V_{ud}| & |V_{us}| & |V_{ub}| \\ |V_{cd}| & |V_{cs}| & |V_{cb}| \\ |V_{td}| & |V_{ts}| & |V_{tb}| \end{pmatrix} \approx \begin{pmatrix} 0.97 & 0.23 & 0.004 \\ 0.23 & 0.97 & 0.04 \\ 0.009 & 0.04 & 0.999 \end{pmatrix}. \tag{2.10}$$

As can be seen from the Eq. 2.10, the diagonal of the matrix is close to unity, which means that the weak interactions within the same generation are the most probable.

The Higgs boson is the only spin 0 boson in SM. Closely related to Higs boson is the process known as the *Higgs mechanism*, which provides masses to elementary particles. It turns out that the minimum of the Higgs potential, also known as electroweak (EW) vacuum, is degenerate, meaning that there are multiple values where it reaches the minimum. Choosing one of these values will cause an phenomenon called *spontaneous symmetry breaking* (SSB). If we also require that after the SSB, the model is invariant under the $SU(2) \otimes U(1)$ local gauge transformation of the electroweak sector, then the Higgs mechanism provides masses for the W and Z bosons. Also the interaction between the non-zero expectation value of the Higgs field and the fermion fields creates a gauge invariant mechanism for generating masses of the fermions.

2.3 State of the electroweak vacuum

In spite SM being the most successful fundamental theory so far, we know that in its current state, it is not complete. There are multiple phenomena that it cannot explain satisfactorily, for example the nature of the dark matter and the pattern of three fermion generations. Correspondingly, there are also different *Beyond Standard Model* (BSM) theories that try to explain the shortcomings of SM, probably the most famous being the supersymmetry of elementary particles.

One of the shortcomings of SM is that it contains many free parameters that have to be measured experimentally. These can be classified into three categories: the coupling constants of the three forces, two parameters specifying the Higgs potential, and dozens of parameters related to how the Higgs field interacts with fermions, including the fermion masses and the mixing angles. Furthermore, the experimentally measured values of these parameters have strong effects on the predictions made by the SM.

One of these free parameters is the mass of the Higgs boson, which is $m_H \approx 125$ GeV. The measured value of m_H makes SM perturbative and predictive when extrapolated to high energies. It turns out that the current measurement of m_H suggests that the EW vacuum is in a *metastable* state [14, 15]. This means that the vacuum is not at the global minimum of the Higgs potential, *i.e.* there exists yet another lower minimum for the potential. This has many interesting consequences. For example this suggests that the current SM might be valid close to the *Planck scale* $M_{\text{Planck}} \approx 10^{19}$ GeV, and therefore new physics could

appear only close that level. Another consequence is that the EW vacuum will eventually collapse into its real minimum. The lifetime of the metastable vacuum can be estimated via the rate of quantum tunneling to the true vacuum minimum predicted by the theory, and the current estimates for the lifetime exceed the age of the universe by orders of magnitude [11].

However, the m_H is not the only parameter affecting to the state of the EW vacuum, as it is also parameterized by the top quark mass, $m_t \approx 172.5$ GeV. This is illustrated in Figure 2.2, from which it is visible that the current estimate for the vacuum state is rather close to the stable section, and that the outer error bars overlap with the metastable-stable border. Actually the known accuracy of

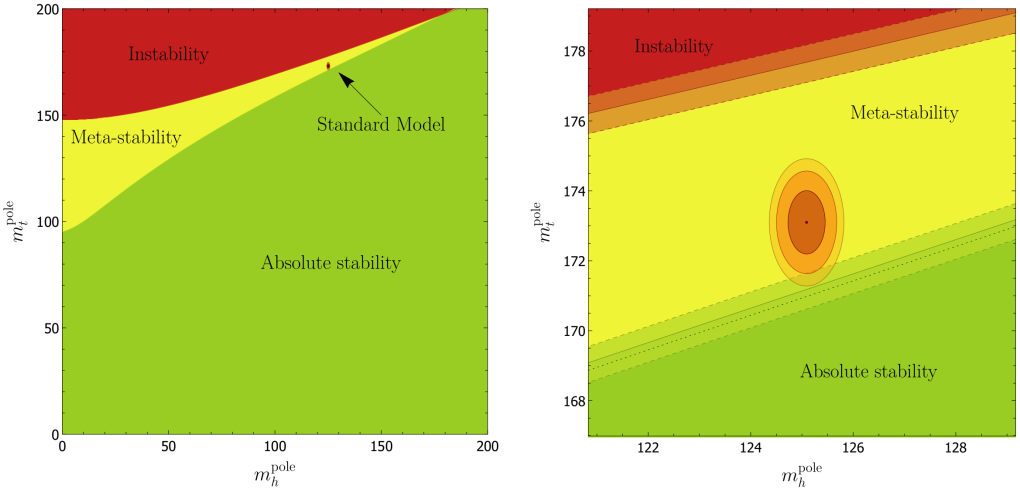

Figure 2.2: Phase diagram of stability of the EW vacuum in terms of m_t^{pole} and m_h^{pole}. The right-hand picture shows the closeup of the SM region, where the ellipses show the 68%, 95% and 99% experimental uncertainties of the masses [16] (modified).

m_t is currently limiting the EW vacuum state predictions, rather than the mass of the Higgs [4], which can be also seen as larger vertical errors compared to the horizontal errors in the 2.2.

One of the main uncertainties to the top quark mass comes from the measurements of b jets. As can be seen from the CKM matrix in Eq. 2.10, the t quark decays almost exclusively into a b quark. In Section 5.4 various phenomena affecting the accuracy of the b jet measurements are discussed and later their effect on the b jet response is evaluated in Chapter 6.

3. Event generation

In this section we are going to take a look at the main features of the modern event generators. We are going to focus on the general purpose Monte Carlo (MC) event generators, which can handle all the simulation steps in high energy particle collisions. The most used general purpose generators at the moment are Pythia 8 (P8) [5] and Herwig 7 (H7) [17].

The objective of event generators is to generate the collision events as precisely as possible. This is not however done in one step, but rather in smaller parts, which together produce approximation of the whole event. The Monte Carlo techniques are used to generate the events according to desired probability distributions, which will produce random final states for the events.

This chapter is divided according to the different phases of the event generation. These four phases are namely hard process, parton showers, hadronisation and decay, and underlying event. The handling of the different phases depend on the energy scale associated with them. Due to the running of the strong coupling, perturbative calculations are possible only at high energies, while the non-perturbative approach has to be used at lower energies.

Special attention is given to the b quark fragmentation, b hadron production fractions, and b and c hadron semileptonic branching ratios, as they play important role in this thesis.

3.1 Hard process

When two protons collide at high energies, one parton from each of the protons take part to the *hard process*. The energies related to the hard process are large, which allows the usage of the perturbation theory in the calculations. The simplest perturbation theory calculations are done at *leading order* (LO), which is the lowest relevant order and the least precise. The higher order calculations yield higher level of precision at the cost of being harder to calculate. These higher orders are named as *next-to-leading order* (NLO), *next-to-next-to-leading order* (NNLO) and so on.

In order to calculate any properties of the hard process, we need to define which parton takes part in the process and which fraction of the protons momentum it carries. This is solved by determining a *parton distribution function* $f_a(x_a, Q^2)$ (PDF), which gives the probability for certain type parton a to carry momentum fraction x at energy scale Q^2.

The PDFs are determined from many experiments by evaluating set of different values of Q^2 and x. However it is not necessary to carry the measurements on the whole Q^2 scale to evaluate the PDFs. The evolution of the PDFs at different energies can be calculated using the DGLAP equations [18, 19, 20, 21]. Figure 3.1 shows an example of two PDFs evaluated at $Q^2 = 10$ GeV^2 and $Q^2 = 10^4$ GeV^2 [22].

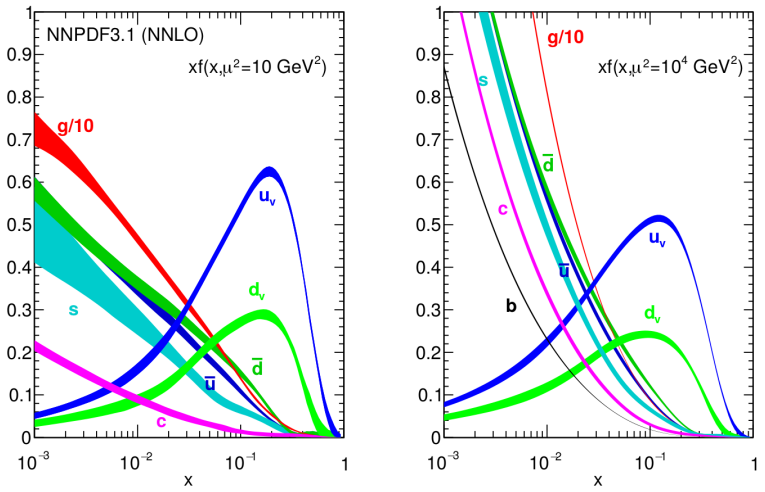

Figure 3.1: PDFs for each parton flavour from NNPDF3.1 NNLO PDF set evaluated at $Q^2 = 10$ GeV^2 (left) and $Q^2 = 10^4$ GeV^2 (right) [22].

The cross section for process with two initial partons a and b to produce final state X is given by

$$\sigma = \sum_{a,b} \int_0^1 dx_a dx_b \int f_a(x_a, Q^2) f_b(x_b, Q^2) d\hat{\sigma}_{ab\to X}(Q^2, \mu_R), \tag{3.1}$$

where $f_a(x_a, Q^2)$ and $f_b(x_b, Q^2)$ are the PDFs, of particles a and b, and $d\hat{\sigma}_{ab\to X}$ is the differential parton level cross section for the process $ab \to X$, which is dependent on the energy scale Q^2 and renormalization scale μ_R [23].

The equation 3.1 can be also written in the form

$$\sigma = \sum_{a,b} \int_0^1 dx_a dx_b \int d\Phi_x f_a(x_a, Q^2) f_b(x_b, Q^2) \frac{1}{2x_a x_b s} |\mathcal{M}_{ab\to X}|^2. \tag{3.2}$$

Now the differential cross section is replaced by the differential phase space element $d\Phi_x$ over X final states. The $|\mathcal{M}_{ab\to X}|^2$ is the *matrix element* (ME) squared, which can be evaluated using various techniques and written as sum over associated Feynman diagrams. The ME represents the probability amplitude for $ab \to X$ process. The $\frac{1}{2x_a x_b s}$ refers to parton flux, and the s in it, denotes the total energy available in the center-of-mass frame for the colliding particles.

3.2 Parton showers

Similar to QED bremsstrahlung radiation, also partons can also particles, namely gluons. As the gluons are themselves colored objects they can further branch and emit more gluons. The branching of gluons from the hard process partons is referred to as *parton shower* [23]. Sometimes, also photons emitted by the partons are called as parton showers.

The gluon can be emitted either from the initial state parton or from the final state parton of the hard process. Radiation from the initial state parton is called *initial state radiation* (ISR), while radiation from the outgoing hard process parton is called *final state radiation* (FSR). Both ISR and FSR approximate the higher-order corrections to the hard process. The state of the art calculations of the hard process are at the moment in NLO and some even in NNLO level, but in order to capture the whole complexity of the collision, the partons showers must be added [11].

The ISR and FSR are added to the event right after the hard process. Each parton is characterized by some evolution parameter Q^2, which determines the time ordering of the branching. For example the Q^2 can be defined as parton's invariant mass-squared, *i.e.* $Q^2 = m^2$. The ISR showers are space-like, with $m^2 = E^2 - |\mathbf{p}|^2 < 0$, leading to negative Q^2 values. In contrast to ISR, the FSR showers are time-like, *i.e.* $m^2 = E^2 - |\mathbf{p}|^2 \geq 0$, and thus the Q^2 values are positive. As the showers are evolved, the values of Q^2 for ISR are increasing as the hard process is approached, whereas the Q^2 values for FSR are decreasing. Thus the ISR can be thought to be evolved backwards, and FSR forwards in time. The shower evolution is then stopped, when some minimum value of Q_0 is reached, which is typically order of 1 GeV [24].

3.3 Hadronisation, fragmentation and decay

At lower energies where strong coupling α_s becomes large, the event generators must rely on phenomenological models to predict how the quarks and gluons will behave. These models are used in the *fragmentation* phase, where the colored partons form color neutral hadrons. These hadrons can be in excited states and still decay further into other particles. The whole process which turns partons into stable hadrons and other particles is called *hadronisation*.

3.3.1 Hadronisation models

The two main hadronisation models are the *string model* and the *cluster model*. The former is used by the P8 and the latter by H7. As this thesis work utilizes the Pythia, the string model is explained in greater detail, as a brief description is only provided for the cluster model.

The cluster model utilizes something called preconfinement property of the QCD. It can be showed that in the evolution scales Q_0 which are much less than the hard process scale Q, the partons are clustered into colourless

groups. Invariant mass distribution of these groups is then independent of the hard process, and only dependent on the evolution scale Q_0 and QCD scale Λ. In the MC simulation this is done by forcing the gluons to split into quark-antiquark pairs at the end of the parton shower. This will create set of light colourless clusters, which can be decayed to on-shell hadrons [23].

The *Lund model* [25, 26] is the most widely used hadronisation model based on strings at the moment. This model can be motivated by considering the strong field between partons. For example if we consider $q\bar{q}$ pair emerging from a hard process, where the quarks move in the opposite directions back-to-back, there will be linearly rising potential between the quarks of the form $V(r) = \kappa r$. The string constant κ have been measured using the hadron mass spectroscopy and it is of the order of $\kappa \approx 1$ GeV/fm [23].

This vast amount of energy stored between the $q\bar{q}$ pair can traced back to the virtual gluons exchanged between the quarks. As the gluons themselves carry the color charge there will attractive forces between these virtual gluons, which will squeeze the color field into a tube between the quarks, as shown in Fig. 3.2. Thus it is reasonable to model the color confinement between the partons as strings.

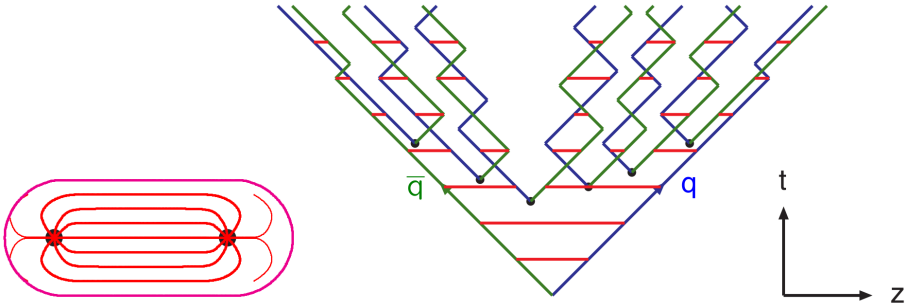

Figure 3.2: The field lines confined to tube-like configuration between $q\bar{q}$ pair (left). Evolution of the string system as $q\bar{q}$ pair move along z-axis (right). The diagonal lines represent quarks and antiquarks, as the horizontal lines depict the string field [23].

If the $q\bar{q}$ pair keep moving into opposite directions, there will be eventually be enough energy stored in the string between them, that a new $q\bar{q}$ pair is created. This is modeled as string breaking, where meson is created from two adjacent string breaks. This phenomenon is illustrated in Fig. 3.2.

3.3.2 Lund fragmentation function

In the simulation, the fragmentation step for light quarks is described by the *Lund fragmentation function* [26], which describes the probability that the daughter hadron will take certain fraction of the total momentum available. It is of the form

$$f(z) \propto \frac{1}{z}(1-z)^a \exp\Big(-\frac{bm_{\mathrm{T}}^2}{z}\Big), \tag{3.3}$$

where z is the fraction of momentum that the daughter hadron will inherit from the parton and m_T is the transverse mass of the hadron defined as $m_{\mathrm{T}}^2 = m_{\mathrm{hadron}}^2 + p_{\mathrm{T,hadron}}^2$. The a and b are free parameters, that are taken to be universal for all quarks, and they are calculated from fits to data.

The Lund fragmentation function needs to be modified for the heavy quarks, *i.e.* charm and bottom. Unlike the light quarks, the heavy quarks are not produced at new spring breaks and do not move along straight lightcone sections. This will lead to exponential suppression of the string and modifications to fragmentation function. The fragmentation function for heavy quarks is called *Lund-Bowler fragmentation function* [27] and it is written as

$$f(z) \propto \frac{1}{z^{1+r_q b m_Q^2}}(1-z)^a \exp\Big(-\frac{b m_{\mathrm{T}}^2}{z}\Big), \tag{3.4}$$

where m_Q is the heavy quarks mass and r_q is shape parameter of the function. Both of these parameters depend on the specific quark type considered.

3.3.3 b hadron production fractions

Different types of b hadrons are not produced in same fractions in the collision events. The fractions also depend on the collision environment, *e.g.* they are different on the $p\bar{p}$ collisions at Tevatron compared to pp collisions at the LHC. Recent data also shows that the fractions depend on the p_{T} of the produced b hadron [28].

It can be assumed that the B^+ and B^0 mesons are produced in the same fractions. Unlike in the charm sector, the relative production rates of B^+ and B^0 are not affected by the strong decays of excited B^{**+} and B^{**0} states and electromagnetic decays of B^{*+} and B^{*0} states. If we also neglect small fraction of the weakly decaying states made with several heavy quarks, we can write following constraints

$$f_u = f_d \quad \text{and} \quad f_u + f_d + f_s + f_{\mathrm{baryon}} = 1, \tag{3.5}$$

where f_u, f_d, f_s, f_{baryon} corresponds to B^+, B^0, B_s^0 and b baryon production fractions.

3.3.4 Particle decays

Due to the hadronisation many new particles are produced. However not all of these particles are stable, so they have to be decayed further. This decaying procedure concerns only the particles that are unstable in the context of the particle detector considered, *i.e.* they have such a lifetime that they might decay before reaching the detector. A typical requirement for stable particles at hadron-colliders is that they satisfy the condition $c\tau \geq 10$ mm, where c is the speed of light and τ is the lifetime of the particle. The most common stable particles are protons $p^{\pm}$, neutrons n, electrons $e^{\pm}$, muons $\mu^{\pm}$, photons γ (mostly

from π^0 decays), charged pions $\pi^{\pm}$, charged kaons $K^{\pm}$, neutral kaons K^0 (mass eigenstate K^0_{L} is stable, while K^0_{S} is not) and lambda Λ^0. These particles make the vast majority of the stable particles, and thus almost all the other particles must be decayed during the simulation.

Particles can decay into several different final states, which are also called as *decay modes*. The probability of some specific decay mode, is referred to as *branching ratio* (BR). Determining which decay mode is used by the MC simulation for each unstable particle is based both on experimental values and theoretically motivated assumptions. For example Pythia 8 uses generally the PDG 2012 tables [29], and decay tables from DELPHI and LHCb for *c*- and *b* hadrons. The particles are decayed isotropically by default, and in addition many of them are then weighted with using generic matrix element, *e.g.* in weak decays. For τ decays, fulls spin correlations are calculated for most of the decay processes. In the case *b* hadrons, the $B^0 - \bar{B}^0$ and $B^0_s - \bar{B}^0_s$ mixing is also simulated before the hadron decay [5].

3.4 Underlying event

The *underlying event* (UE) represents all the other activity in the event which is not directly related to the hard parton-parton interaction and its associated ISR and FSR. The major effect to the UE is believed to come from *multiple parton interactions* (MPI). In MPI, more than one parton-parton interaction occurs in the collision of the two hadrons. These interactions can be modeled either perturbatively or nonperturbatively. In the perturbative approach, the MPI is modelled mainly as t-channel gluon exchange, as the nonperturbative model is based on exchange of *pomerons*, which can be viewed as colour-singlet fluctuations consisting of two gluons [11].

If there happens more than one hard interaction in the same hadron collision, these will typically produce back-to-back dijets. However soft interactions which will not produce observable jets are much more likely, and can affect significantly the event topology. The soft MPI contribute to UE by increasing the total amount of scattered energy and by inducing more color exchange between the partons, which will lead to production of more particles during the hadronisation. The MPI can also radiate gluons and create parton showers, which will increase the complexity of the MPI modelling.

It is also possible that multiple proton-proton collisions occur in the same time window in which proton bunches collide. As distinction from the specific proton-proton interaction under the study, these overlapping additional collisions are known as *pileup* (PU).

3.5 Tunes

The event generators contain number of free parameters, which have to be set before running the simulation. While default values exist for these parameters in

different programs, usually one can achieve better correspondence between data and simulation by *tuning* the MC model. The tuning means in practice that the free parameters are defined according to some certain experimental data and/or more accurate theoretical calculations. The given set of parameters is called as a *tune*.

In practice, the tunes can define any number of free parameters. However, most commonly the tunes modify the parameters related to UE, hadronisation, ISR, FSR and the choice of the PDF. The latest tune used at the CMS at the moment is the `CP5` tune [30].

4. The Large Hadron collider and the CMS experiment

Particle accelerators are the main tools for studying the elementary particles. By accelerating and colliding different high energy particles, mainly protons and electrons, SM can be put under stringent tests. Interactions between high energy particles will manifest themselves as production of variety of different hadrons, leptons and bosons, which in turn can be observed with particle detectors.

In this chapter we take a brief look to the *Large Hadron Collider* (LHC) and one its main experiments, *Compact Muon Solenoid* (CMS). At end of this chapter, detector simulations are reviewed briefly. The LHC section is mostly based on references [31, 32], of which the former is the full technical report of the LHC.

4.1 The Large Hadron Collider

The LHC is the largest man-made particle collider in the world. It was built by European Organization for Nuclear Research (CERN), which today control its operations and upgrades. The LHC is located in a tunnel which lies between 45 m and 170 m beneath the France-Switzerland border near the city of Geneva. The total length of the LHC is approximately 27 km and it consists of eight straight and eight arc sections creating ring-like shape.

In order to reach energies at TeV-scale, LHC utilizes the old particle accelerators at the same site to pre-accelerate the particles before they enter the LHC itself. Schematic picture of LHC with its pre-accelerators is shown in Fig. 4.1. Inside the LHC, the particles travel inside two beam pipes in opposite directions. The pipes are kept in ultra high vacuum and the particles inside the pipes are guided and focused using superconducting magnets. Dipole magnets are used to bend the trajectory of the particles, while quadrupole magnets focus the particle beam. In order to create the superconducting state, the magnets are cooled down to 1.9 K using liquid helium.

The two pipes cross each other at four collision points, prior which the particles are squeezed closer together with magnets to enhance the probability of the collisions. Around these four collision points, are the four biggest particle detectors CMS, ATLAS, ALICE and LHCb. All the detectors are specialized for studying certain phenomena: CMS and ATLAS are general detectors focused on

SM and BSM theories, ALICE studies mainly quark-gluon plasma and conditions of the early universe through heavy-ion collisions, while LHCb is focused on the CP violation and matter-antimatter asymmetry in the universe through observations of b hadrons.

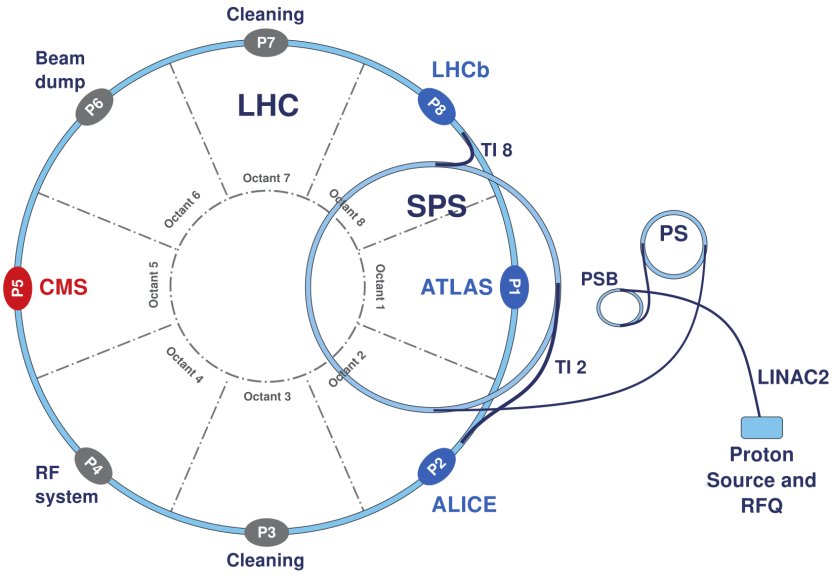

Figure 4.1: Schematic picture of LHC, pre-accelerators and location of different detectors around the beam pipe (not in scale) [33].

Most of the collision experiments at the LHC are done by colliding protons. In order to enhance the collision probability, the protons are collided in bunches, which contain approximately $1.15{\cdot}10^{15}$ protons each. The bunches are accelerated close to the speed of light and multiple bunches circulate at the same time inside the beam pipe separated by 25 ns interval. In addition to protons, also heavy ions, such as lead, are collided in the LHC. Heavy-ion collisions are used to study QCD matter at extreme temperatures and energy densities [34].

The operation of LHC is divided into active collision periods called *Runs* and longer repair and upgrade periods called *Long Shutdowns*. The operation of LHC started with Run 1, which lasted from 2010 to 2012. The initial center-of mass energy of $\sqrt{s} = 7$ GeV was increased to $\sqrt{s} = 8$ GeV in 2012, while the total integrated luminosity of pp collisions produced during the Run 1 was about 30 fb^{-1} [35]. The Run 2 started in 2015, with upgrades done during the Long Shutdown 1, LHC now reached energy of $\sqrt{s} = 13$ GeV. Data taking ended in 2018, with total integrated luminosity of 160 fb^{-1}. Currently the LHC is under Long Shutdown 2, while the Run 3 is planned to begin in 2022 with increased energy of 14 TeV [36].

As LHC has only collected about 6% of the planned total integrated luminosity of 3000 fb^{-1} [37], it has already made many major discoveries. So far the biggest finding has been the detection of the Higgs boson in 2012 [1, 2]. The high energies of LHC have also enabled finding many new hadrons, including rare penta- and tetraquark states [38]. Besides new discoveries, LHC has had

important role in precise measurements of the SM parameters and pushing limits of the BSM theory parameters.

4.2 The CMS detector

The following section about the CMS is divided into four parts. First, the structure of the detector is reviewed, based on the detailed description of the detector of the reference [39]. After this, it is explained how the individual particles are reconstructed and identified using the Particle Flow algorithm [40]. In the last two subsections, the jet clustering and calibration at the CMS are reviewed.

4.2.1 Detector design

The CMS is general purpose particle detector located underground, near Cessy in France, 100 m below the surface at the collision Point 5. The CMS detector has an overall diameter of 15 m, length of 22 m and it weighs 14 000 tonnes. The schematic picture of the detector is shown in Fig. 4.2.

The central property of the CMS is a superconducting solenoid with radius of 6 m and capability to provide 3.8 T magnetic field. The CMS itself consists of multiple sub-detectors, each having its own special purpose in the particle detection. Moving outwards from the beam pipe, the CMS consists of a silicon pixel and strip tracker, a lead tungstate crystal electromagnetic calorimeter (ECAL) and a brass and scintillator hadron calorimeter (HCAL), which all are located inside the solenoid.

The muon detection system is located outside the solenoid and consists of gas-ionization chambers embedded in the iron yoke. Both ECAL and HCAL consists of a barrel and two endcap detectors. Also forward calorimeters can be found at each end of the CMS detector, which extend the overall pseudorapidity coverage of the CMS.

The particle collisions produce massive amounts of data, which all cannot be saved for the data analysis. Thus a *trigger system* is utilized to decide which events to preserve for further analysis. At the CMS the events are selected using using two-tiered trigger system. The first level trigger (L1) uses information from calorimeters and muons system to select events within a fixed latency of 4 μs at a rate around 100 kHz [42]. After L1, the high level trigger is applied (HLT). The HLT reduces the event rate to around 1 kHz before the data is stored. It runs a highly optimized and fast version of the full event reconstruction, which allows usage of more complicated event cuts [43].

4.2.2 Particle Flow algorithm

The *Particle Flow* (PF) algorithm aims to identify the final state particles and their properties produced in the *pp* collision by combining the data from all the sub-detector systems [40]. By using the distinct features of the particles,

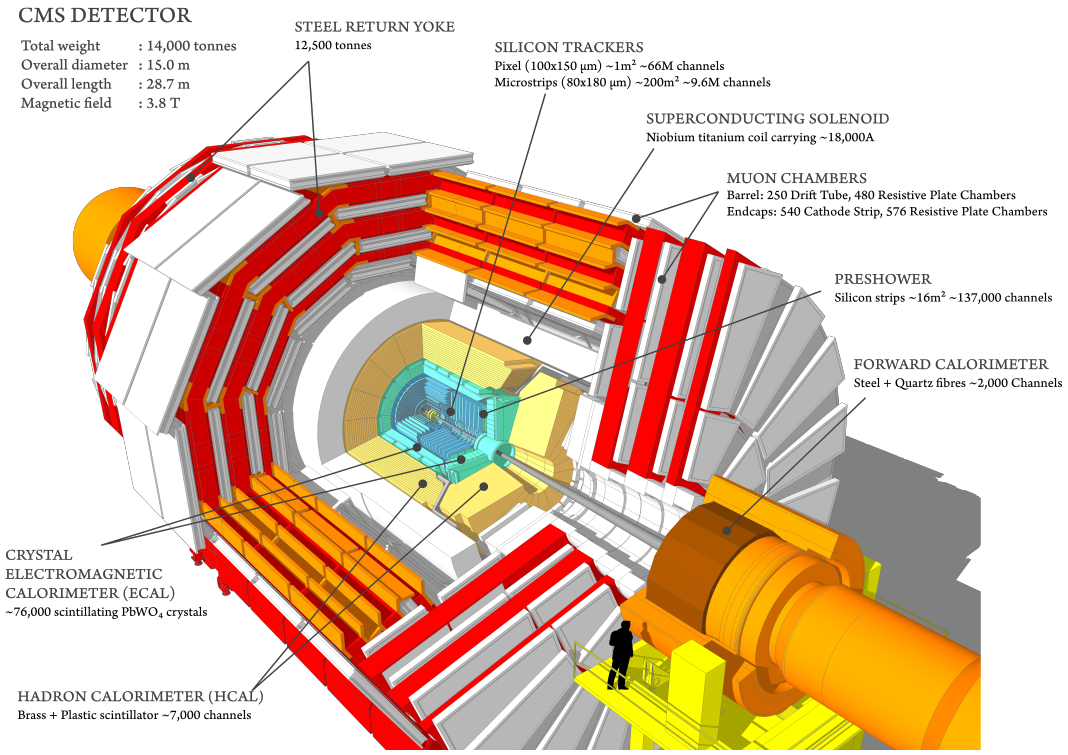

Figure 4.2: Schematic picture of CMS detector showing the different sub-detectors [41].

PF identifies particles as photons, electrons, muons, charged hadrons or neutral hadrons.

After collision, particles first encounter the tracker. Both charged-particle tracks and decay vertices are reconstructed using the signals from the tracker. As the superconducting solenoid creates the strong magnetic field, charged-particles fly in bent trajectories, from which their electric charge and momentum can be measured.

After the tracker, particles enter the ECAL in which the electrons and photons are absorbed. These are detected as signals in multiple adjacent calorimeter cells, also referred to as *clusters*, from which the direction and the energy of the particles can be calculated. Also hadronic showers from charged and neutral hadrons can produce moderate signals in the ECAL.

The charged and neutral hadrons are finally absorbed totally in the HCAL. The muons go through the tracker, ECAL and HCAL as minimally ionising particles and they are detected by the muon detector system located outside the solenoid. None of the sub-detectors can detect the neutrinos, which traverse through the CMS with no interactions.

Next, we go through how the individual particle types (photons, electrons, muons, charged hadrons and neutral hadrons) are determined by the PF algorithm. As photons are not electrically charged, they do not leave a track in the tracker and their trajectories do not bend in the magnetic field. As a result pho-

tons can be identified by ECAL energy clusters without a track. Photons can originate from π^0 decays, which are typically reconstructed as a single ECAL cluster.

Electrons, on the other hand, are associated with bent charged track and ECAL clusters linked to them. Additionally, photons emitted from the electrons due to the interactions with the detector material, can be used for electron identification. Muons can be detected by charged track linked to hits in the muon detector system.

The hadron identification depends on whether the hadron is charged or neutral. Charged hadrons produce charged particle track linked to ECAL and HCAL energy clusters. It is also demanded that the track is not associated before neither to electron or muon.

Lastly, the neutral hadrons can be identified by HCAL clusters, which are not linked to any of the charged hadron tracks. Neutral hadrons can be also reconstructed from the charged hadron clusters in the ECAL and HCAL, if the total calibrated energy of the clusters is bigger than the sum of charged particle momenta linked to it from the tracker.

The energies of the different particle types can be also calculated using the sub-detector information. The photon energy is estimated from the measured ECAL energy deposit. The energy of the electron is reconstructed by combining the tracker and ECAL estimates, with the energy of the bremsstrahlung photons associated with the track. The muon energies are determined using the tracker information if the transverse momentum of the muon is less than 200 GeV. Otherwise the energy is approximated by the best track fit combined to the energy deposits from the muon detector system.

The charged hadron energy is obtained by combining the track momentum with the ECAL and HCAL energy deposits linked to that particular track. The calorimeter energy deposits are recalibrated for the nonlinear response of ECAL and HCAL. Similarly the neutral hadron energy is determined from the recalibrated HCAL and ECAL energy clusters. The directions of charged particles are determined from the direction of the track at the vertex, while the directions of neutral particles are calculated from the location of the associated calorimeter energy cluster.

When all the particles and their energies are reconstructed, PF estimates the *missing transverse momentum* $\vec{p}_{\mathrm{T}}^{\,\mathrm{miss}}$, also referred to as missing missing transverse energy (MET) for historical reasons. The MET reveals the presence of particles, like neutrinos, which do not interact with the sub-detectors.

The raw MET vector is the negative vector sum of the all reconstructed particle momenta

$$\vec{p}_{\mathrm{T}}^{\,\mathrm{miss,raw}} = -\sum_{i=1}^{N_{\mathrm{particles}}} \vec{p}_{\mathrm{T},i}. \tag{4.1}$$

The corrected $\vec{p}_{\mathrm{T}}^{\,\mathrm{miss}}$ is calculated by utilizing calibration, which propagates the corrections made to p_{T} of the jets to the MET [44].

4.2.3 Jet clustering

In order to study jets and compare results involving jets we need a consistent definition for them, *i.e.* some well defined set of rules that project group of individual particles into jets. There are many ways to define such a rules and the different set of rules are called *jet clustering algorithms*.

There are several requirements for good jet clustering algorithm. The algorithm should be collinear and infrared safe (IRC safety). This requires that the jet properties cannot change if one of the constituent particles is replaced by a group of collinear particles carrying the same total momentum, and that the properties of the jet do not change if infinitely many soft particles are added to the event. Using IRC safe algorithms ensure that the results of the clustering can be compared to the jet cross sections and other properties calculated using the QCD perturbation theory [45].

The two main classes of the algorithms are sequential recombination algorithms and cone algorithms. Historically cone algorithms were favoured in collision experiments as they were easier to implement. As a drawback, most of them are IRC unsafe. Thus majority of the experiments at the LHC rely on the sequential recombination algorithms which are IRC safe and available nowadays as computationally fast implementations in the FastJet software package [46].

The three most popular sequential clustering algorithms are Cambridge/Aachen [47], k_t [48] and anti-k_t [49]. In general, the sequential recombination algorithms use bottom-up approach in the jet clustering. First a distance measure is declared, then pair of particles which are closest to each other are identified according to that measure. The pair is then recombined after which the next pair of closest particles is searched. This iteration process is continued until some stopping criterion is reached.

The CMS utilizes the anti-k_t in the jet clustering with distance parameter $R = 0.4$. The anti-k_t algorithm clusters the hard particles first and produces circular cone shaped jets. As it is also insensitive to UE and PU, and a robust implementation is available in the FastJet, its usage is favoured at LHC.

4.2.4 Jet calibration

The measured and reconstructed jet energy do not match with the energy of the jet at particle level. The detectors bias the measurement with noise and limited precision of their of calibration. Also additional bias to energy measurement is coming from the pileup events. For these reasons jets need to be calibrated.

In order to get the jets to the correct jet energy scale (JES), series of jet energy corrections (JEC) are applied to take into account all the different sources of error. The diagram in figure 4.3 shows the order of successive JEC applied to MC simulations and data at the CMS. The calibration methods discussed in this section are developed for the Run 1. Detailed description of these methods can be found for $\sqrt{s} = 7$ TeV in Ref. [50] and for $\sqrt{s} = 8$ TeV in Ref [3].

The first corrections to be applied are the pileup offset corrections (also

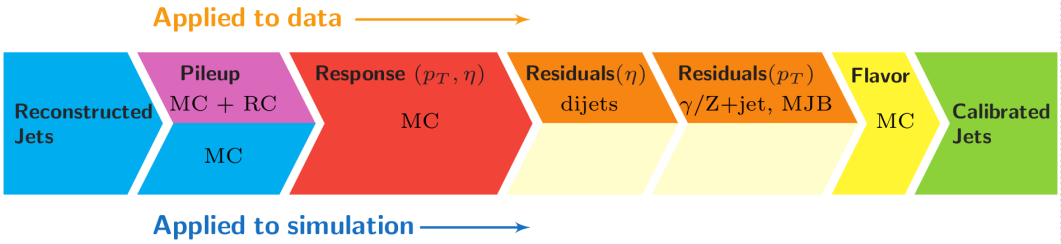

Figure 4.3: Successive stages of JEC applied to data and MC simulations at the CMS [3].

referred to as L1 corrections). The pileup contribution coming from the same beam crossing as the measured jets are called in-time pileup, as the pileup coming from previous or subsequent beam crossing are called out-of-time pileup. The effect on JES by both of these are tried to minimized with suitable corrections.

The effect on particle level pileup offset is evaluated by MC simulation, where same event is reconstructed with and without pileup background, after which the corresponding jets are matched. The correction factor between MC and data is calculated using random cone method on the zero-bias events. This means calculating the average p_{T} of the reconstructed particles in randomly placed cone using sample which does not contain hard interactions and is selected by random triggers during the bunch crossings.

After the pileup corrections, the simulated response corrections (L2L3) are applied to the jets. L2L3 corrections aim to correct on average the energy disparity between the reconstructed jets and the simulated jets. These corrections are derived using the full CMS detector simulation based on GEANT4 with QCD multijet events generated with Pythia8 with `CUETP8M1` tune. The goal is to take into account the biases coming from different detectors, *e.g.* non-linear calorimeter response.

To compute the response corrections, the simulation level jets are matched with the reconstructed jets. The condition for matching requires that the jets are within $\Delta\mathrm{R} < 0.2$ from each other. The upper limit for $\Delta\mathrm{R}$ is half of the cone size used in reconstruction. The simulated jet response is then calculated in $p_{\mathrm{T}}^{\mathrm{gen}}$ and $|\eta^{\mathrm{jet}}|$ bins as $\langle p_{\mathrm{T}}^{\mathrm{reco}} \rangle / \langle p_{\mathrm{T}}^{\mathrm{gen}} \rangle$, where $p_{\mathrm{T}}^{\mathrm{reco}}$ and $p_{\mathrm{T}}^{\mathrm{gen}}$ are reconstructed and simulated jet p_{T}, respectively.

After L2L3 corrections, the residual corrections (L2L3Res) are applied. The residual corrections aim to correct the remaining small differences in the jet response between data and simulations. The correction factors are calculated as ratio of jet responses between data and the simulated samples.

The jet responses are evaluated from data and simulations by two methods: p_{T} balance and MPF (missing transverse momentum projection fraction). In the p_{T}-balance method, the jet response is defined as

$$R_{p_{\mathrm{T}}^{\mathrm{bal}}} = \frac{p_{\mathrm{T}}^{\mathrm{jet}}}{p_{\mathrm{T}}^{\mathrm{ref}}}. \tag{4.2}$$

In this method, the response is evaluated by comparing the jet p_{T} directly to

p_{T} of the reference object. The p_{T}-balance method is based on the selection of the γ+jet, Z+jet, dijet or multijet events, where the γ, Z, central jet or recoil system can be measured with better accuracy than than the jet.

The MPF method takes different route in the response evaluation. It exploits the properties of the event topologies, where MET is not present in the hard process. It is defined as

$$R_{\mathrm{MPF}} = 1 + \frac{\vec{p}_{\mathrm{T}}^{\,\mathrm{miss}} \cdot \vec{p}_{\mathrm{T}}^{\,\mathrm{ref}}}{(\vec{p}_{\mathrm{T}}^{\,\mathrm{ref}})^2}. \tag{4.3}$$

The MPF method takes account the whole hadronic part of the event, not just the measured jet, when comparing its p_{T} to the recoiling reference object p_{T}.

Both of these methods of calculating the response can be biased by additional jets in the event. This will cause p_{T} imbalance between the probe jet compared to the reference object. As the additional jets will depend on the studied event topology and are not correlated with the jet energy response, this effect has to be taken into account in the energy corrections. This additional jet activity is quantified with variable α, which is ratio between the first subleading jet of the event, divided by the typical p_{T} scale of the event. The corrections are extrapolated to the value of $\alpha = 0$, so that the effect of additional jets is removed.

The L2 residual corrections are calculated using dijet events, and thus they are also known as relative residual corrections. The reference jet is required to be in the $|\eta| < 1.3$ barrel region of the detector, to obtain good precision. The probe jet has unconstrained pseudorapidity, which leads to corrections normalized to the central region of the detector. These corrections are derived in the p_{T} and $|\eta|$ bins of the probe jet.

The L3 residual corrections use Z+jet, photon+jet and multijet events, and are also called as absolute residual corrections. In these events, the JES is evaluated by comparing the jet p_{T} to precisely measured reference objects. These objects can be electrons or muons decayed from the Z boson or photons. In the multijet events, the response is evaluated with multijet balance method (MJB) [51] or using MPF. The MJB is based on comparing the high p_{T} jet in the barrel region to two or more lower p_{T} jets.

The different event types are used to cover different p_{T} ranges. Roughly speaking, the Z+jet events are used at low, photon+jet at medium and multijet at high p_{T} ranges. However this division is not absolute and different event types typically partially overlap with each other. The absolute jet p_{T} scale is then fitted simultaneously to all three event types. Fig. 4.4 shows the data/MC ratio for 2016 data, with different event type comparison and the global fit.

The last correction in the JES calculation are the flavour corrections. As the jets originating from different flavours (uds, c, b, g) behave differently during the hadronisation, flavour dependent correction factors need to be derived. These are calculated using samples of jets tagged with truth-level flavour information in simulation.

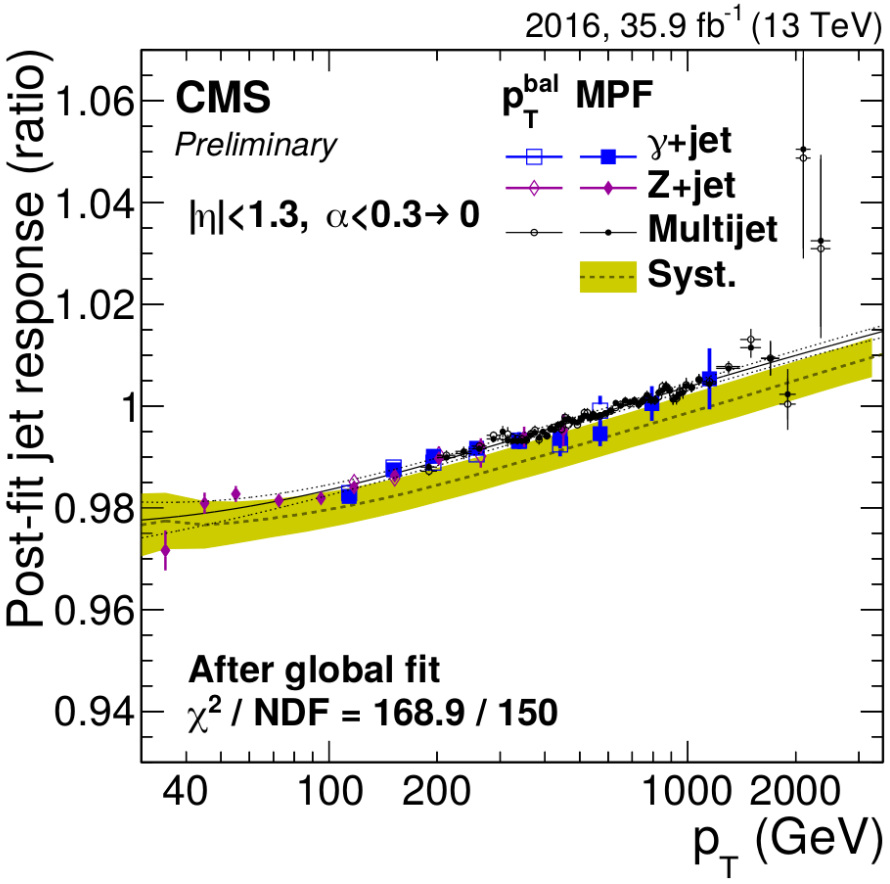

Figure 4.4: Comparison of the response ratio of data and simulated Z+jet, γ+jet and multijet samples. The yellow band indicates absolute scale uncertainty that is centered around the luminosity-weighted average of JEC per run. The post-fit jet response refers to method where the points are adjusted by their associated uncertainties to fit a common smooth function [52].

4.2.5 b jet tagging

Precision measurements of SM parameters and searches of BSM signal rely on the efficient identification of the b jets. For example, as explained in Section 2.3, the stability of the vacuum is dependent on the top quark mass, which in turn is dependent on the b jet measurement.

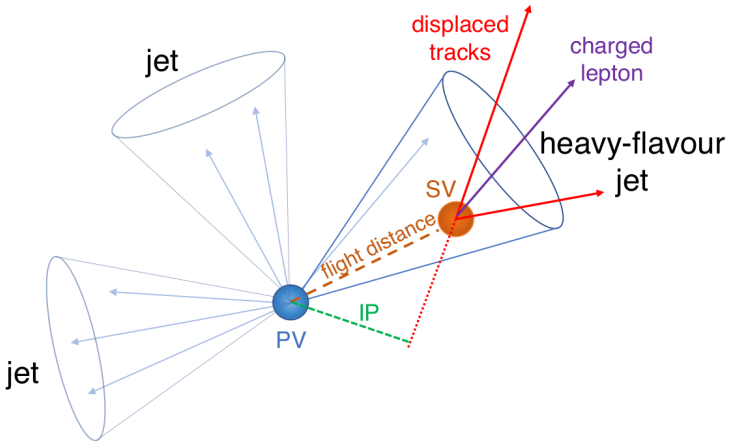

Figure 4.5: Illustration of the event topology of the heavy-flavour jet. The secondary vertex is displaced from the primary interaction vertex due to the long lifetime of the b and c hadrons [53].

The identification of jets originating from b quarks is called *b-tagging* [53]. The b-tagging is based on the relatively long lifetime of b hadrons, which is order of 1.5 ps. This leads to creation of *secondary vertex* (SV) which is typically from few mm to one cm away from the primary vertex. The displaced tracks originating from the SV, is thus sign of a b jet. Another sign of b jet is the presence of a charged lepton, as approximately in 20% of the events, electron or muon is produced in the decay chain of b hadron. Figure 4.5 illustrates the event topology of heavy-flavour jet event (b or c jet).

There are several b-tagging algorithms in use at the CMS. These include CSVv2 and DeepCSV, which both are based on the training of neural network specialized on the identification of b jets. The b jet identification efficiency for $\sqrt{s} = 13$ TeV data is evaluated with precision of a few percent at $30 < p_{\mathrm{T}} < 300$ GeV and about 5% in the range of $500 < p_{\mathrm{T}} < 1000$ GeV.

4.3 Detector simulation

In order to compare the data simulated with the MC generators to experimental data, it has to be run through a *detector simulation*. The detector simulation, as its name suggests, simulates the workings of the detector and the interactions between the detector material and the particles.

There are various approaches to simulating the detectors. One option is to simulate the full geometry of the detector and its materials, which can be

done by using the GEANT4 toolkit [54]. The GEANT4, which can simulate the electromagnetic and hadronic interactions between of particles and various materials, allows building complex detector geometries and to monitor the particle tracking.

The complete detector simulation at CMS utilizing it is called *CMS Full Simulation* [55]. In addition to the detector hits, obtained from input particles ran through the GEANT4, the Full Simulation also considers simulated hits from pileup and noise models, before producing the output.

As simulating the whole detector is computationally very time consuming, a faster version of CMS detector simulation has been created, namely the *CMS FastSim* [56]. Rather than simulating the whole detector very precisely, in the FastSim material effects are taken into account only at hit level. This allows approximately 100 times faster simulation and about 20 times faster simulation with a reconstruction, but still comparable accuracy.

Another approach is to use parametric detector simulation like DELPHES framework [57]. Instead of simulating the whole detector, DELPHES uses the parameterized detector responses, and information about efficiencies and mistagging factors. This allows very fast simulation, but on contrary cannot achieve high level of accuracy. The detector simulation used in this thesis is also based on parameterized model. It omits modelling resolution effects for single particles, which increases the statistical precision for jet response effects [58]. This model is presented in more detail in Section 5.3.

5. Methods

In this chapter, first the used event generation environment and the studied event type with its selection criteria are presented. After this, the parameterized detector simulation and reconstruction program used for the studies is described. The end of this chapter will then focus on the rescaling method and the various features that were rescaled.

5.1 Event generation environment

In this work Pythia 8.244 (P8) [5] was used to generate the events, with the `CP5` [30] tune. We used the `NNPDF31_nnlo_as_0118` PDF from the `LHAPDF 6.2.0` [59]. The jet clustering was performed using FastJet 3.3.0 [46] utilizing anti-k_t algorithm with a jet distance parameter $R = 0.4$. The simulated particles were stored as ROOT tuples using ROOT 6.18.04 [60]. These tuples contain all the final state particles, clustered jets, partons originating from the hard process and heavy hadrons containing c and b quarks before their decay to other particles.

5.2 Event topology and selection

As relevant to the low p_{T}-range JES studies, Z+b jet events were chosen as the subject of the thesis. In Fig. 5.1 there are Feynman diagrams showing the possible LO hard process event types.

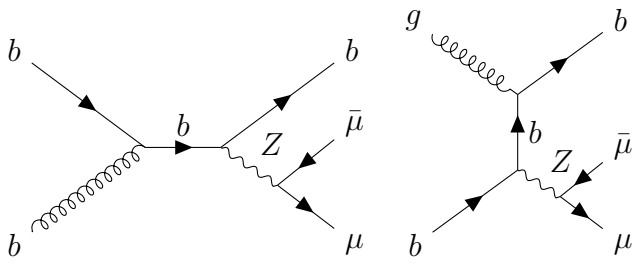

Figure 5.1: The LO Feynman diagrams of the $Z(\rightarrow \mu\mu)$+b events in the s-channel and t-channel, respectively.

The event generation of Z+jet events of P8 were modified so that the outgoing parton from hard process was always b quark and that the Z boson

would always decay into $\mu\bar{\mu}$ pair. The $\mu\bar{\mu}$ pair serves as reference object in the MPF and p_{T}-balance responses discussed in section 4.2.4. The muons can be measured and reconstructed with good precision which makes them as ideal reference objects.

Several event cuts were also applied to the events after the generation and jet clustering. The additional jet activity was reduced by demanding following α-cut for the events

$$\alpha = \frac{p_{\mathrm{T}}^{2^{\mathrm{nd}}\mathrm{jet}}}{p_{\mathrm{T}}^{Z}} < 0.3, \tag{5.1}$$

where $p_{\mathrm{T}}^{2^{\mathrm{nd}}\mathrm{jet}}$ refers to jet with second highest p_{T} in the event and $p_{\mathrm{T}}^{\mathrm{Z}}$ is the reconstructed Z boson p_{T} from the $\mu\bar{\mu}$ pair. In the future analyses the α cut is going to change to value $\alpha < 1$. It is also required that the angle between the reconstructed Z and the leading p_{T} jet of the event is

$$|\Delta\phi| > 2.8 \text{ rad}. \tag{5.2}$$

This ensures that the reference object and leading jet are traversing almost in opposite directions. Thus it is sensible to assume that the reconstructed Z boson can be used to estimate the true transverse momentum of the leading jet.

There are also following additional p_{T} and η cuts for the leading jet and the reference object

$$p_{\mathrm{T}}^{1^{\mathrm{st}}\mathrm{jet}} > 15 \text{ GeV}, \quad |\eta_{1^{\mathrm{st}}\mathrm{jet}}| < 1.3 \tag{5.3}$$

$$p_{\mathrm{T}}^{Z} > 15 \text{ GeV}, \quad |\eta_{\mathrm{Z}}| < 2.5 \tag{5.4}$$

$$p_{\mathrm{T}}^{\mu} > 15 \text{ GeV}, \quad |\eta_{\mu}| < 2.3. \tag{5.5}$$

These cuts ensure that the leading jet is in the barrel region of the detector and the reference object is within the tracker acceptance, where they can be measured more precisely. Also minimum p_{T} value is set for both leading jet and for the reference object as this is the lowest jet p_{T} used in the CMS analyses. Jets with even lower p_{T} are harder to reconstruct and measure precisely.

Finally, large MET values and invariant mass of Z boson were restricted as

$$p_{\mathrm{T}}^{\mathrm{miss}} < 0.9 \cdot p_{\mathrm{T}}^{\mathrm{Z}} \tag{5.6}$$

$$70 < m_Z < 110 \text{ GeV}. \tag{5.7}$$

As the mass of the Z boson is $m_Z \approx 91.2$ GeV, the mass limits ensure that the measured leptons are from resonant Drell-Yan process, *i.e.* from virtual Z boson rather than from virtual photon. The mass cuts also prevent cases where hard FSR is emitted from one of the leptons decayed from the Z boson.

5.3 Parameterized detector simulation

The detector simulation and the particle reconstruction in this thesis work is based on parameterized model of the CMS PF reconstruction, called *toyPF*,

which was developed as part of the Master's thesis [58], and implements the method originally developed by D0 [61, 62], to study flavour-dependent JES. The toyPF replicates few main features of the PF reconstruction and uses single-particle responses and efficiencies, which allows very fast simulation and sufficient precision to study jet response uncertainties.

Most importantly, toyPF omits single-particle resolution effects, which improves statistical precision for jet quantities significantly, as uncertainties scale linearly with jet resolution, but only as square-root of effective number of events. The method ensures a low statistical uncertainty reaching sub-permille statistical precision with reasonably sized sample and avoids statistical limitations of FullSim samples for systematic studies. ToyPF also conserves the information of the flavour of the jet during the reconstruction, thus it is well suited for the b jet studies in this thesis.

As input toyPF takes a list of stable particles and clustered jets. Then detector response is assigned to all particles according to its type and spatial orientation. After this, calorimeters are replicated by collecting the energy deposits of the particles in 2D-histograms covering $-5.2 < \eta < 5.2$ and $0 < \phi < 2\pi$ following the HCAL granularity, which is 0.087×0.087. These cells represent the calorimeter clusters in the PF reconstruction. The position of particle hits in the calorimeters are calculated for charged particles by taking into account the magnetic field inside the detector, which will create curved particle tracks. The model also takes account the tracking efficiency, which represents the probability that the particle is reconstructed correctly as charged hadron. Finally, the reconstructed charged particles are linked to the calorimeter cells, so that the total energy deposit of the cell can be calculated using the track momenta and calorimeter energy deposits.

5.4 Rescaling

The b jet energy scale (bJES) was studied using simulated P8 samples and toyPF detector simulation. Certain features of SM in P8 are assumed to be modelled imprecisely, and results in unwanted calorimetric energy depositions and extra tracks, which suffer poor MC prediction. In order to fix this, we can rescale the individual events so that averaged P8 results coincide with the experimental ones. Similar method was used by ATLAS in their top quark mass measurement [63], where they scaled b quark fragmentation, hadron production and decays to match MC sample to the experimental data.

The events can be rescaled with *scaling factors* c_i defined as

$$c_i(p_{\mathrm{T}}) = \frac{n_{i,\mathrm{exp}}(p_{\mathrm{T}})}{n_{i,\mathrm{MC}}(p_{\mathrm{T}})}, \tag{5.8}$$

where n_{exp} and n_{MC} refer to event counts of some variable in a bin i, for experimentally measured data and data acquired from MC simulations, respectively. Generally the scaling factors are p_{T}-dependent and evaluated per p_{T}-bin in this

thesis, unless stated otherwise. The scale factors are only calculated from and applied to according to the highest p_{T} b hadron of the event. Using the b hadron p_{T} instead of b jet p_{T} was motivated by comparing the p_{T} dependency of branching ratios of semileptonic b hadron decays. As shown in Figure 5.2, the BRs are more or less p_{T}-independent when they are plotted as a function of b hadron p_{T}, as the BRs plotted as function of jet p_{T} are clearly dependent on the p_{T}. This effect is probably due to the lack of neutrinos in the generator level jets, which causes jets to be placed in lower bins, as the p_{T} of neutrinos is not taken into account.

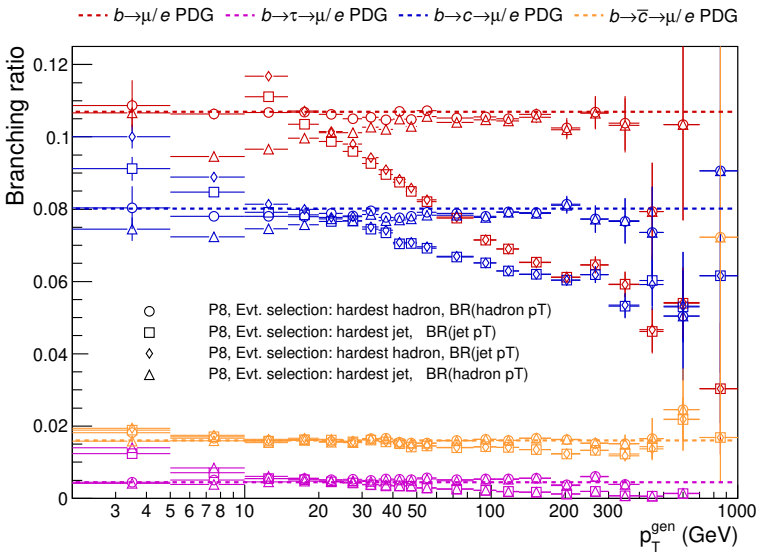

Figure 5.2: Effect on the semileptonic b hadron BRs, when the scaled event is selected either by the hardest jet or the hardest b hadron of the event, and whether the BRs are functions of jet p_{T} or hadron p_{T}.

The scaling factors are applied to generated P8 sample before running the detector simulation. In practice, the scaling is done by multiplying the event weights with the c_i, and filling the histograms with this new event weight. In order to estimate the effect of the scaling procedure, detector simulation is also run with unscaled sample. After this the bJES can be determined from both samples and the two results be compared.

To ensure that the rescaling would not change the properties of the original sample, it was required that the total number events cannot change due to the scaling. This is the main reason that scaling factors were calculated separately for each p_{T} bin. It was also demanded that scaled features would not produce unphysical results, meaning that, *e.g.* production fractions sum up to one after the scaling.

Next all the 4 different scaled features are represented one by one in more detail. These features are semileptonic BR of b hadrons (two variations), b-to-c hadron BR, b hadron production fractions and tuning of the Lund-Bowler fragmentation function parameterization. All the PDG values used in the following

section are also listed in Appendix B, where the corresponding PDG decay channel numbers are also listed.

5.4.1 b hadron semileptonic branching ratio

The b hadron semileptonic BR was studied with two different approaches. One approach considered decay of admixture of different types of b hadrons and the other inspected decay of certain b hadron types. In other words, the former approach examines inclusive decay modes, as the latter one examines exclusive modes. In both cases decay chain leading to production of electron or muon was assumed to be sign of semileptonic decay, as fully leptonic decays of b hadrons are very rare and thus negligible. The b hadrons were given a tag corresponding to the different decay channels at the generator level, which could be used later to calculate the branching ratios.

The reference values for admixture of b hadrons are from the newest edition of PDG [11]. The admixture contains $B^{\pm}$, B^0 and B_s^0 mesons and b baryons. Four different decay channels are considered, which are listed in Table 5.1 with the corresponding PDG values and P8 values calculated from a simulated sample. For the $b \to \tau \to \mu/e$ mode, the BR is calculated using PDG value of BR($b \to \tau^+\nu_\tau$ + anything) and the of possible subsequent decays of τ via channels BR($\tau^- \to \mu^-\bar{\nu}_\mu\nu_\tau$), BR($\tau^- \to \mu^-\bar{\nu}_\mu\nu_\tau\gamma$), BR($\tau^- \to e^-\bar{\nu}_e\nu_\tau$) and BR($\tau^- \to e^-\bar{\nu}_e\nu_\tau\gamma$).

Table 5.1: Semileptonic branching ratios of inclusive b hadron admixture modes. In $b \to c$ and $b \to \bar{c}$ modes, the b and c refer to a quark inside the hadron. Note that b corresponds to B antimeson or b baryon, as B mesons contain a $\bar{b}$ quark, while b baryons contain a b quark. Situation is simpler with c hadrons, as all c hadrons contain a c quark and all antihadrons a $\bar{c}$ quark [11].

Decay mode	PDG(%)	P8 avg.(%)
$b \to \mu/e$	10.69 ± 0.22	10.60 ± 0.01
$b \to \tau \to \mu/e$	0.45 ± 0.04	0.52 ± 0.01
$b \to c \to \mu/e$	8.02 ± 0.19	7.82 ± 0.01
$b \to \bar{c} \to \mu/e$	1.6 ± 0.5	1.6 ± 0.1

Figure 5.3 shows the initial semileptonic branching ratios and PDG values. The scale factors for each channel are solved numerically using ROOT for each p_{T}-bin. The system of equations from which the scaling factors were solved can be found in Appendix A. The right-hand side plot on Figure 5.3 shows the corresponding scale factors.

Similar scaling method was used by ATLAS for their studies [63]. However, they only studied the semileptonic decays to muons, and did not take the electron into account. In that study, the scale factors were also p_{T}-independent, thus only one scale factor was applied to whole p_{T}-range. The average P8 BR values shown in Table 5.1 correspond fairly good to the values reported in the ATLAS study.

Table 5.2: Semileptonic branching ratios for individual b hadron types [11].

Hadron	PDG(%)	P8 avg.(%)
$B^{\pm} \to \mu/e$	10.99 ± 0.28	11.29 ± 0.01
$B^{0} \to \mu/e$	10.33 ± 0.28	10.44 ± 0.01
$B^{0}_{s} \to \mu/e$	9.6 ± 0.8	9.3 ± 0.1
$\Lambda^{0}_{b} \to \mu/e$	10.9 ± 2.2	7.7 ± 0.1

The small differences may be due to the inclusion of the electrons in our studies or the different P8 tune used by ATLAS, which might affect to the b fragmentation. Also the reference PDG values differ as the ATLAS study used the 2018 PDG version [64].

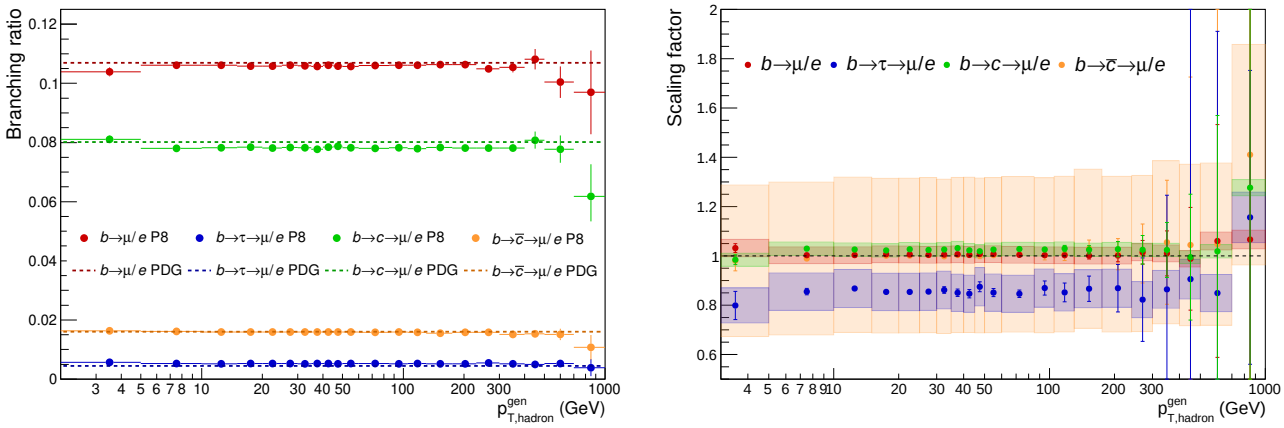

Figure 5.3: Initial P8 branching ratios for semileptonic admixture decay channels and the corresponding PDG values [11] (left). The associated scaling factors with the statistical error shown by the error bars and systematic error by the colored band (right).

The other method to rescale the semileptonic b hadron decays calculates the scaling factors individually for $B^{\pm}$, B^{0} and B^{0}_{s} mesons and Λ^{0}_{b} baryons. Again the reference values are from the latest PDG version [11] and shown in Table 5.2. This time the scale factors can be solved easily analytically, see Appendix A for the derivation and exact formulation of the scale factors. Figure 5.4 shows the branching fractions measured from the P8 with the PDG values, and the corresponding scaling factors.

Similar method is utilized in the official CMS analyses [65], in which the systematic uncertainties related to top quark measurements are estimated. The rescaling is done for the branching ratios of $B^{\pm}$, B^{0} and B^{0}_{s} mesons and Λ^{0}_{b} baryons in order to match them with PDG values. This time only the direct decays of leptons from b hadrons were considered and chain decays via c hadrons and τ leptons were omitted. In this study, the scaling factors were also p_{T}-independent and obtained by averaging the branching ratios from the whole

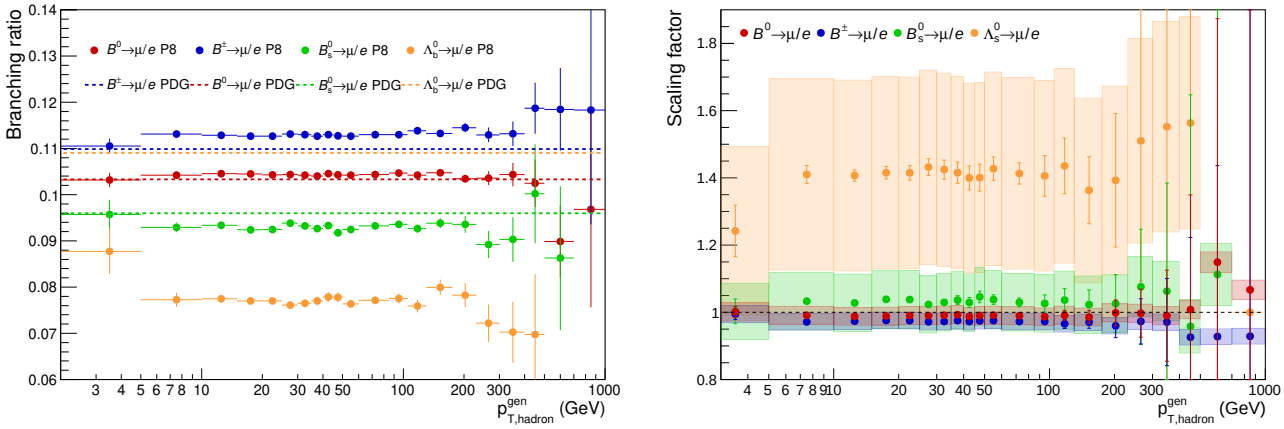

Figure 5.4: Initial P8 branching ratios for individual b hadron semileptonic decay channels and the PDG values [11] (left). The associated scaling factors with the statistical error shown by the error bars and systematic error by the colored band (right).

p_{T}-range. The average P8 BR and PDG values in Table 5.2 correspond exactly to the values used in Ref. [65].

5.4.2 b-to-c hadron inclusive branching ratios

Inclusive branching ratios of b-to-c hadron decays were rescaled to the latest PDG values [11]. Again the b hadrons were tagged at the generator level according to the possible subsequent decays to c hadrons. Also the possibility of B meson decaying into two D mesons were taken into account. Table 5.3 shows the decay channels chosen to this work.

All other channels except $B_s^0 \to D_s^-$ considered decays to either meson or antimeson states. As the BR of this channel exceeds 100%, when varied up by one sigma uncertainty, the upper uncertainty limit was restricted to 7%, *i.e.* $\mathrm{BR}(B_s^0 \to D_s^-) = 93^{+7}_{-25}\%$. For the decay modes $B^0 \to D^+$ and $B^0 \to D_s^-$ only the upper bound was given for BRs at 90% confidence level. These upper bounds were propagated to the systematical uncertainties of the decay modes $B^0 \to D^-$ and $B^0 \to D_s^+$ without affecting to the nominal values of these modes. Symmetric uncertainties were used for these modes.

The decay mode $B_s^0 \to D_s^-$ was excluded from the rescaling of the central values, as the uncertainties of the mode are so large. However these uncertainties were propagated to systematic uncertainties of the final results. Again the scale factors were solved numerically using ROOT and the solved system of equations is presented in Appendix A. Figure 5.5 shows the inclusive branching ratios calculated from the P8 with the corresponding PDG reference values. Figure 5.6 shows the solved scaling factors for each p_{T}-bin.

Table 5.3: b-to-c hadron inclusive branching ratios [11]. [a] Consists of decay modes BR($B^0 \to D^-$) = 36.9 ± 3.3% and BR($B^0 \to D^+$) < 3.9% which is given at 90% confidence level. [b] Consists of decay modes BR($B^0 \to D^+$) = 10.3±2.1% and BR($B^0 \to D^-$) < 2.6% which is given at 90% confidence level. [c] The upper uncertainty limit of this decay channel was restricted to 7% in the calculations in order to keep the the results physically reasonable.

Hadron	PDG(%)	P8 avg.(%)
$B^0 \to D^0/\overline{D}^0$	55.5 ± 3.2	46.03 ± 0.1
$B^0 \to D^+/D^-$ [a]	36.9 ± 5.1	50.8 ± 0.1
$B^0 \to D_s^+/D_s^-$ [b]	10.3 ± 3.3	12.9 ± 0.1
$B^+ \to D^0/\overline{D}^0$	87.6 ± 4.1	83.8 ± 0.1
$B^+ \to D^+/D^-$	12.4 ± 1.3	12.0 ± 0.1
$B^+ \to D_s^+/D_s^-$	9.0 ± 1.4	13.7 ± 0.1
$B_s^0 \to D_s^-$ [c]	93 ± 25	82 ± 1

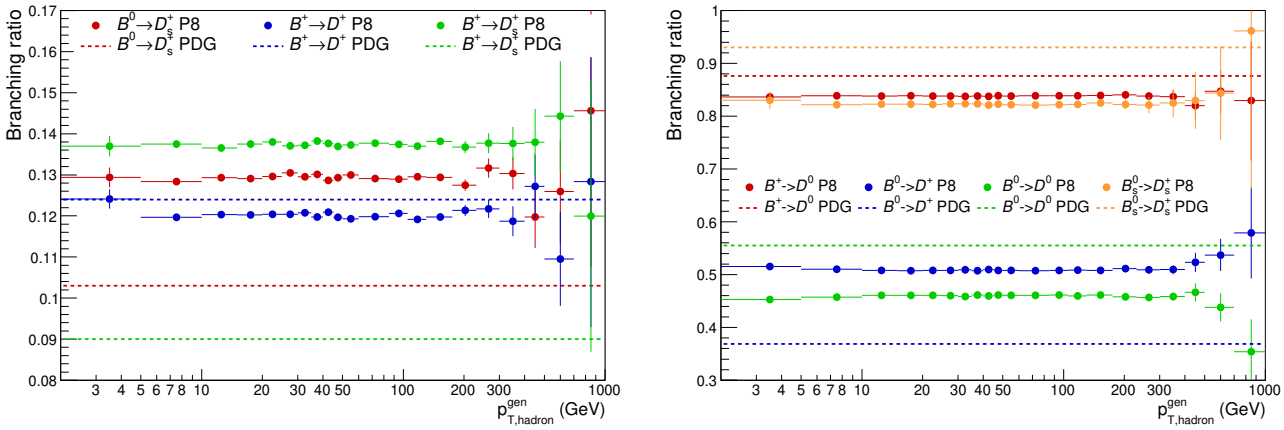

Figure 5.5: Initial P8 branching ratios for variety of b hadron to c hadron decay channels and the PDG values. The presented decay modes include also the D antimeson states except in the case of B_s^0 decays [11].

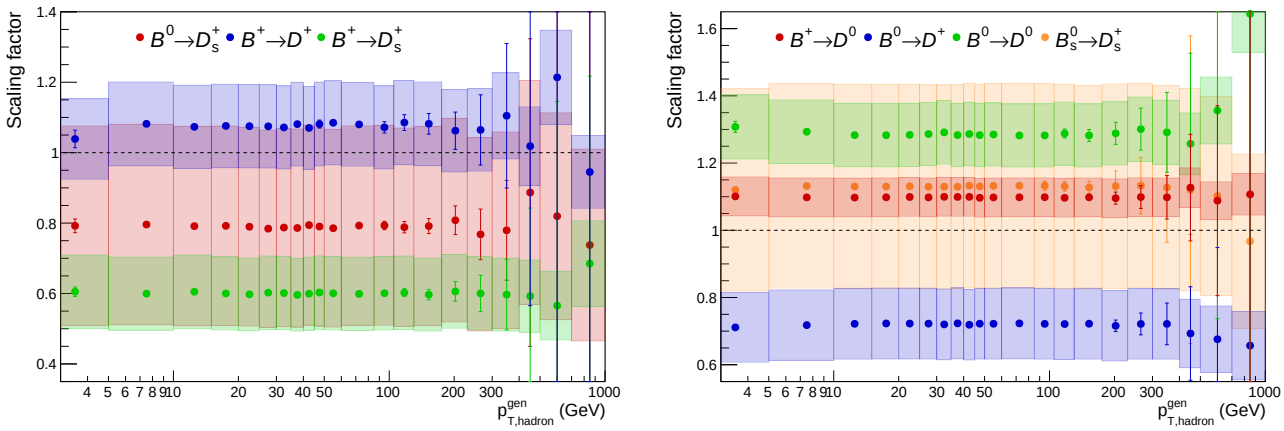

Figure 5.6: Numerically solved scaling factors for the b hadron to c hadron rescaling. The left-hand side plot contains the scaling factors for channels presented in the plot on the left in Fig. 5.5 and similarly the right-hand side contains the scaling factors for the rightmost plot from the same figure. The presented decay modes include also the D antimeson states except in the case of B_0^s decays.

5.4.3 b hadron production fractions

For the production fraction rescaling, combination of two different reference values were used. At low p_{T} we use the latest LHCb measurement of the production fractions of $\bar{B}_s^0$ and Λ_b^0 hadrons normalized to the sum of B^- and $\bar{B}^0$ using the 13 TeV collision data corresponding to integrated luminosity of 1.67 fb^{-1} [66]. The study was done for b hadrons with p_{T} in range from 4 to 25 GeV and pseudorapidity constrained from 2 to 5. The study found no dependency between the ratio of fractions and pseudorapidity. However, it was found out that the ratio of Λ_b^0 had strong and ratio of $\bar{B}_s^0$ mild dependence on p_{T} of the b hadron.

Following the notation presented in Section 3.3.3, the normalized fractions of $\bar{B}_s^0$ and Λ_b^0 are denoted as $\frac{f_s}{f_u+f_d}$ and $\frac{f_{\Lambda_b}}{f_u+f_d}$. The LHCb data and the fits for the normalized production fractions as function of b hadron p_{T} are shown in Figure 5.7.

To estimate the normalized fractions at higher energies, the PDG [11] values for production fractions of B^-, $\bar{B}^0$ and $\bar{B}_s^0$ were used. The PDG contains only the production fraction of b baryons in general, and not separately for Λ_b^0 . Thus the production fraction of Λ_b^0 from all the b baryons was first estimated with P8 using 20M fully b-enriched sample, yielding $f_{\Lambda_b^0}/f_{\mathrm{baryon}} \approx (78.2 \pm 0.1)\%$, after which the b baryon production was scaled with this value. The normalized fractions were then calculated by using the definitions $\frac{f_s}{f_u+f_d}$ and $\frac{f_{\Lambda_b}}{f_u+f_d}$. The PDG values used for the calculations are shown in Table 5.4.

In Figure 5.8 both LHCb and PDG reference data are plotted with the corresponding values calculated with P8. The PDG data points are placed at 200 GeV as the results of Ref. [66] suggests that the fractions are p_{T}-dependent only at low energies. The actual scaling factors are calculated using the values

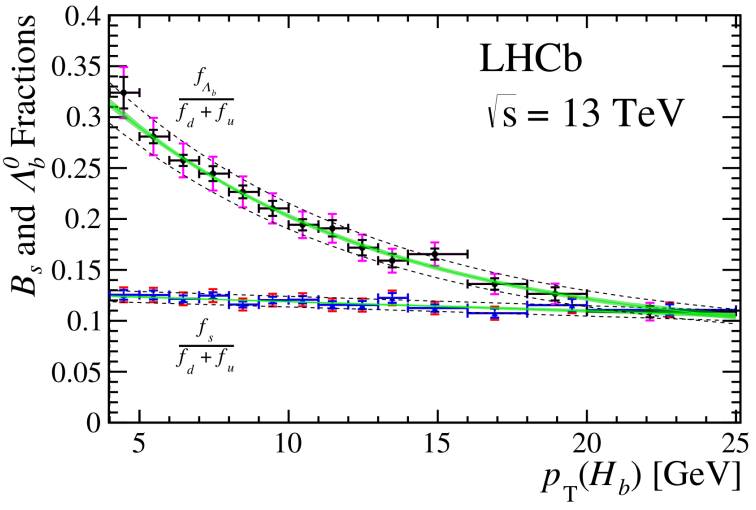

Figure 5.7: Data for $f_s/(f_u + f_d)$ and $f_{\Lambda_b^0}/(f_u + f_d)$ are indicated by black solid circles and blue triangles respectively. The smaller error bars show the the combined statistical and systematic uncertainties, while the larger ones show the global systematics added in quadrature. The fits to data are shown as solid lines [66].

Table 5.4: b hadron production fractions [11]. [a] The $f_{\Lambda_b^0}$ is calculated using the PDG value of f_{baryon} and the P8 estimate for Λ_b^0 fraction out of all b baryons. See Section 5.4.3 for more details.

Hadron	PDG(%)	P8 avg.(%)
B^+ (f_u)	40.8 ± 0.7	42.9 ± 0.1
B^0 (f_d)	40.8 ± 0.7	42.9 ± 0.1
B_s^0 (f_s)	10.0 ± 0.8	9.5 ± 0.1
b baryon (f_{baryon})	8.4 ± 1.1	4.7 ± 0.1
Λ_b^0 $(f_{\Lambda_b^0})$ [a]	6.6 ± 0.9	3.7 ± 0.1

from the fits made to reference data points. A constant and exponential function are fitted to reference values of $\frac{f_s}{f_u+f_d}$ and $\frac{f_{\Lambda_b}}{f_u+f_d}$, respectively. The functions are fitted using ROOT and are parameterized as

$$g_{B_s^0} = 0.117857 \tag{5.9}$$

$$g_{\Lambda_b} = 8.02613 \cdot 10^{-2} + \exp(-0.980386 - 0.110891 \cdot x). \tag{5.10}$$

Especially the g_{Λ_b} was chosen to take into account the p_T-dependency on the low energy range and the fact that the fractions are assumed to be constant at higher energies. The fit function for B_s^0 used by Ref. [66] seems suitable only for the energy range used in the study, *i.e.* from 4 to 25 GeV, as it keeps decreasing as p_T increases. There is no physical justification to assume that fractions should decrease linearly at higher energies, eventually reaching negative values, thus a

different fit function was chosen for this study. A different fit function was also chosen for Λ_b^0 , as we wanted to take account high p_{T} data point at 200 GeV too. See Appendix C for the original fit functions used by the LHCb.

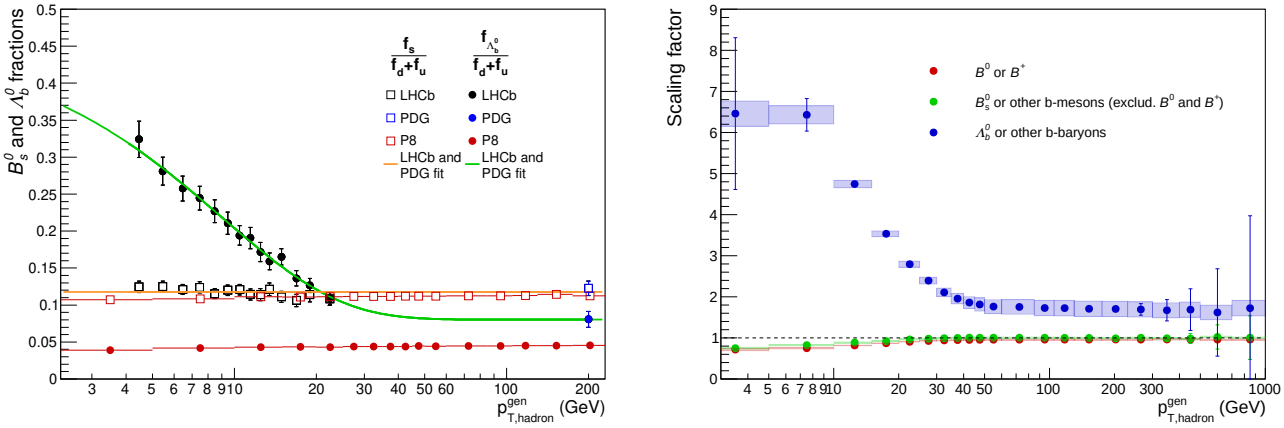

Figure 5.8: The reference values for the b production fraction scaling from LHCb [66] and PDG [11] with the associated fits and the values calculated from P8 (left). The associated scaling factors with the statistical error are shown by the error bars and systematic error by the colored band (right).

The scale factors were calculated using fit functions $g_{B_s^0}$ and g_{Λ_b} values at the middle of each p_{T}-bin. Again it was demanded that the number of events is unchanged after the scaling. The exact derivation of the scale factors is found in Appendix A. Figure 5.8 shows the derived scale factors. It was also checked that the production fractions did not change at the generator level between between different η regions used by the LHCb compared to this thesis, *i.e.* $2 < |\eta| < 5$ and $1.3 < |\eta|$, respectively. Plot of the comparison can be found in Appendix C.

It was assumed that the same scale factor is applicable for B^+ and B^0 mesons, as their production fractions are expected to be the same. It was also assumed that the production of other b baryons would be similar to Λ_b^0, so all the b baryons were scaled with same scale factor. The production fraction of all the other B mesons except B^0, B^+ and B_s^0 is very small. Thus they could be neglected, but for the sake of consistency and more straightforward implementation of the code, they were scaled with the same scale factor as B_s^0.

Production fraction scaling was done also by ATLAS in their top mass measurement [63]. However, they assumed that the production fractions were p_{T}-independent and they used global averages for the rescaling, which mixed the measurements from LEP and Tevatron. The latest version of the summary of the HFLAV group results suggests that these global averages should not be used [28]. This for the reason that the fractions seem to be dependent on production environment, thus the fractions measured at one collider are not directly applicable for the other colliders. It is also unknown how the p_{T} dependency varies between different event types as the LHCb studied inclusive sample. Thus it is not certain that similar effect is present in $Z + b$ and $t\bar{t}$ events.

5.4.4 b quark fragmentation

The b quark fragmentation was modified by changing the parameterization of the Lund-Bowler fragmentation function in P8. The value of the r_B-parameter was changed from the default $r_b = 0.855$ value to the reference value $r_b = 0.858 \pm 0.037(\text{stat}) \pm 0.031(\text{syst})$, which was recently measured by the CMS [67]. The analysis used charm mesons produced inside b jets from $t\bar{t}$ pair decays at $\sqrt{s} = 13$ TeV with integrated luminosity of 35.9 fb^{-1}.

As mentioned in Section 3.3.2, the r_q parameters determine the shape of the fragmentation function. Now as we are studying the b jets, we assume that changing the value of r_b will also have effect on its fragmentation. We can study this by observing the p_{T}-ratio of the b hadron and the jet that it is included in. The variable x_B describes this p_{T}-ratio and is defined as

$$x_B = \frac{p_T^{B\text{-hadron}}}{p_T^{\text{jet}}}. \tag{5.11}$$

Fig 5.9 from the CMS study mentioned before shows the x_B distribution when evaluated with the newly measured r_B and with the default P8 value.

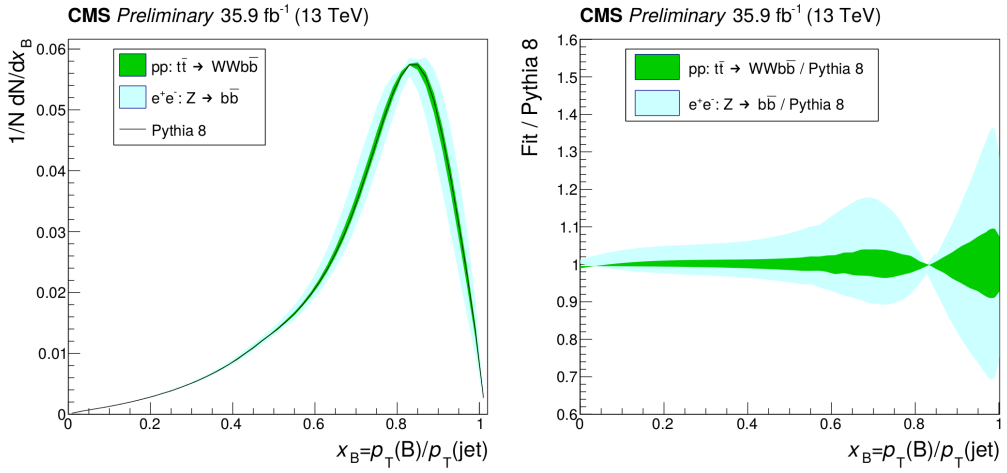

Figure 5.9: The Lund-Bowler fragmentation function for bottom quarks as function of $x_B = p_{\text{T},B\text{-hadron}}/p_{\text{T,jet}}$ (left). The green and light blue bands shows the results of the analysis, while the black line corresponds to Pythia 8 default value. The derived x_B distribution divided by the P8 values are shown on the right-hand side plot [67].

To obtain the scaling factors, two P8 samples were produces: one with the default r_B value and one with the newly measured reference value. The effect of changing the r_B on the x_B distribution can be seen in Figure 5.10. The scaling factors are derived by dividing the x_B values bin-wise from the generated default and reference distributions. As the Ref. [67] used the `CUETP8M1` tune for their studies, the reference distribution is also generated with that tune, while the default distribution is generated with `CP5` tune. The scaling factors are shown in Figure 5.10.

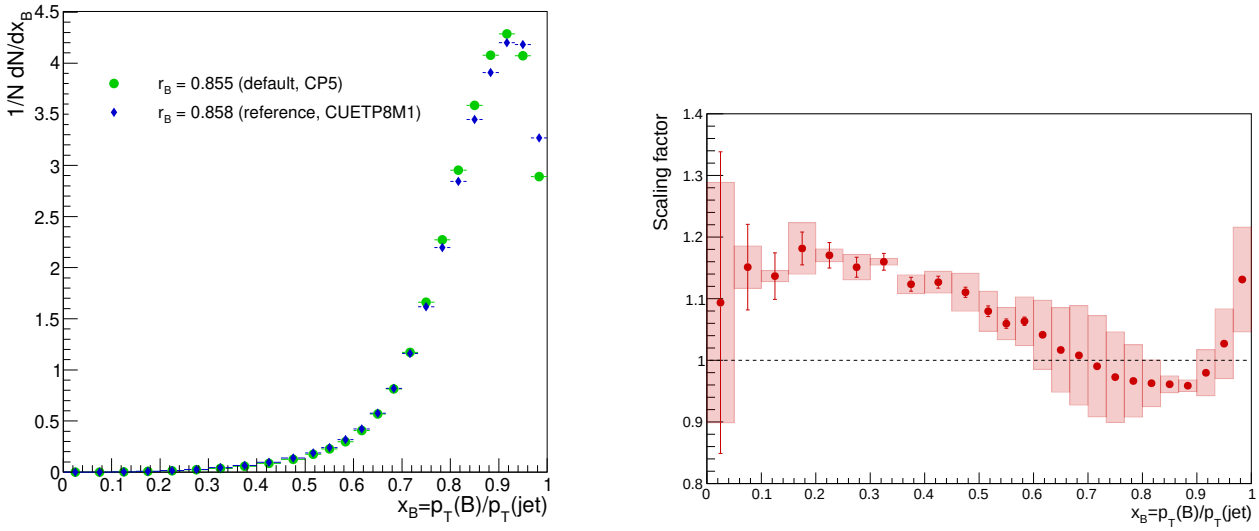

Figure 5.10: Two x_b distributions generated using different values of r_B-parameters. The reference values are from the latest CMS experiment measurement [67] (left). The scaling factors derived by calculating the ratio of the reference and default x_B distributions (right). The statistical uncertainty is shown by the error bars and systematic uncertainty with the colored band.

5.4.5 Uncertainty estimation

The systematic uncertainties arising from the scaling factors were evaluated by varying all of them up and down within one standard deviation, after which the jet responses were calculated again with the varied scale factors. These varied jet responses are later referred plainly to as up and down variations, while the unvaried response is called as central variation. The statistical uncertainties arising from the scaling factors were propagated directly to the statistical uncertainties of the responses, which especially affects to the precision at the high p_{T}.

Two methods were used to estimate the uncertainty of the total response shift. If the up and down variations of all the rescaling methods are added in quadrature, it will result in symmetric uncertainty around the total shift calculated with central scaling factor values. In this method we assume that all the scaling methods work perfectly and the reference values chosen are fully applicable for this analysis.

The other total uncertainty estimate takes into account the contributions from all the different rescaling methods more equally. Again the up and down variations are added in quadrature but this time one of them is always excluded from the calculations. After all the methods have been excluded and the total uncertainties are calculated for each of these variants, the values giving the biggest uncertainties are chosen and will represent the final total uncertainty.

The latter approach is used in Chapter 6, in which the total uncertainty is shown on the background of all the plots in Sections 6.1.5 and 6.2. The variants

of the plots of Sections 6.1.5 and 6.2 using the former method are presented in Appendix E.

6. Results

First results of the rescaling are presented separately for each of the rescaling methods in Section 6.1. After this, in Section 6.2, the total effect on the response is presented when all the scaling methods are applied.

As there are two different scaling methods for semileptonic BRs of b hadrons, only one of them can be used at a time for the calculation of the total effect on the responses. In this chapter we chose the scaling of individual b hadron types instead of the admixture of them, as this method is also utilized in the CMS official top quark mass measurements [65]. The alternative versions of the plots of Sections 6.1.5 and 6.2 using the inclusive model instead can be found in Appendix F.

In this chapter, the detected shifts are inspected mainly on the range $p_{\mathrm{T}} < 300$ GeV, as the statistical uncertainty of the results will increase rapidly after that. However, the results are shown up to 1000 GeV, as the b hadron production fraction and especially b quark fragmentation rescaling yield good accuracy also at high p_{T}.

6.1 Effects of individual rescaling methods on the jet response

In all the following plots in this section, the response differences are shown for MC truth, MPF and p_{T}-balance responses. In the plots, the central variation refers to result obtained by scaling with the central values of the scaling factors. The up and down variations are calculated as discussed in Section 5.4.5, and they represent the systematic uncertainties related to the reference values. The x-axis is labeled as '$p_{\mathrm{T,jet}}^{\mathrm{gen}}$ or $p_{\mathrm{T,tag}}^{\mathrm{reco}}$', as the MC truth response is given as function of $p_{\mathrm{T,jet}}^{\mathrm{gen}}$, whereas the MPF and p_{T}-balance responses as are functions of $p_{\mathrm{T,tag}}^{\mathrm{reco}}$. The differences between different jet responses were discussed in Section 4.2.4.

In order to study how neutrinos affect the responses, an alternative set of plots was created where the neutrinos were clustered to MC jets. These plots can be found in Appendix G. If clustering of the neutrinos to the MC truth response reproduces the MPF and p_{T}-balance shifts, it is a sign of neutrinos creating the difference between the MC and experimental responses.

6.1.1 b hadron semileptonic branching ratios

The effect of scaling the semileptonic BRs of b hadron admixture is shown in Figure 6.1. As can be seen, the effect on the MC truth response is very small. However, the up and down variations of the MPF and p_{T}-balance are significant and about 0.2% at middle p_{T}-range. The effect decreases at $p_{\mathrm{T}} < 50$ GeV. As explained above, the differences between MC truth response and experimental methods are due to the neutrino production.

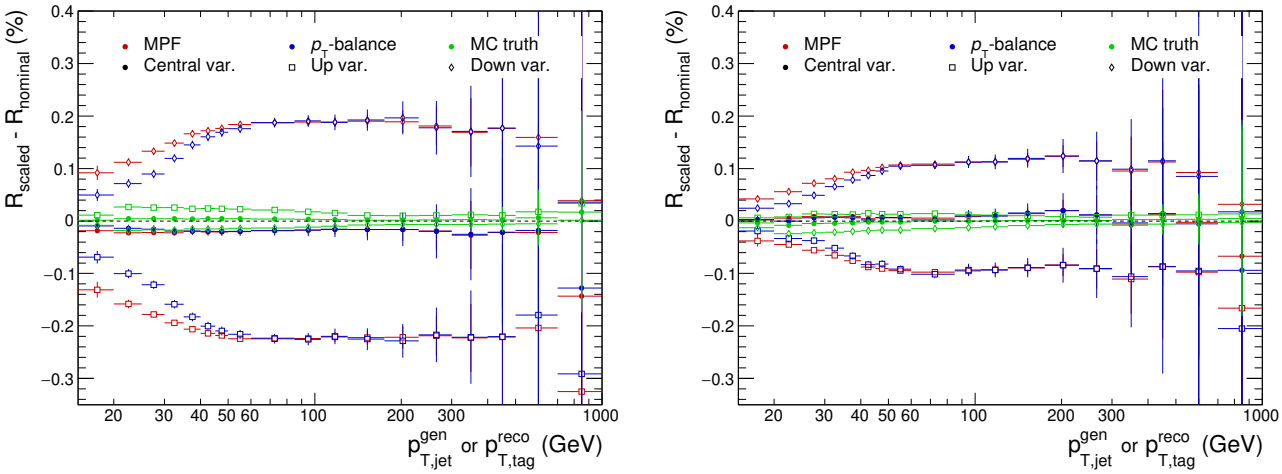

Figure 6.1: Change in the b jet responses, when the semileptonic branching ratios of inclusive b hadron admixture decay channels were rescaled (left). Change in the b jet responses, when semileptonic branching ratios of individual b hadron types were rescaled (right).

The effect of the other semileptonic BR scaling, namely the scaling of different b hadron types, is also shown in Figure 6.1. Again the the effect on the MC truth response seems to be negligible and bigger effect is seen on the variations of the other response types. The up and down variations show about 0.1% effect on the MPF and p_{T}-balance responses, and again the effect decreases at $p_{\mathrm{T}} < 50$ GeV. This method yields smaller shifts in the up and down variations as it takes account only the prompt semileptonic decays. It does not rescale decay chains $b \to c \to \mu/e$, $b \to \bar{c} \to \mu/e$ and $b \to \tau \to \mu/e$, which are included in the inclusive rescaling method.

6.1.2 b-to-c hadron inclusive branching ratios

The effect of the b-to-c hadron inclusive BR scaling is shown in Figure 6.2. This time there is a clear difference between the central variations of MC truth and experimental responses. The central value is shifted almost +0.05% at $p_{\mathrm{T}} = 30$ GeV, after which the shift decreases until disappearing around $p_{\mathrm{T}} = 200$ GeV. The up and down variations show about 0.1% effect on both side of the shift.

For MPF and p_{T}-balance the shift is a bit bigger and to the opposite direction. At $p_{\mathrm{T}} = 300$ GeV the shift is slightly larger than 0.05%, and it decreases as the p_{T} decreases until disappearing at $p_{\mathrm{T}} = 30$ GeV. The variations show shift of 0.2% on the both side of the central variation.

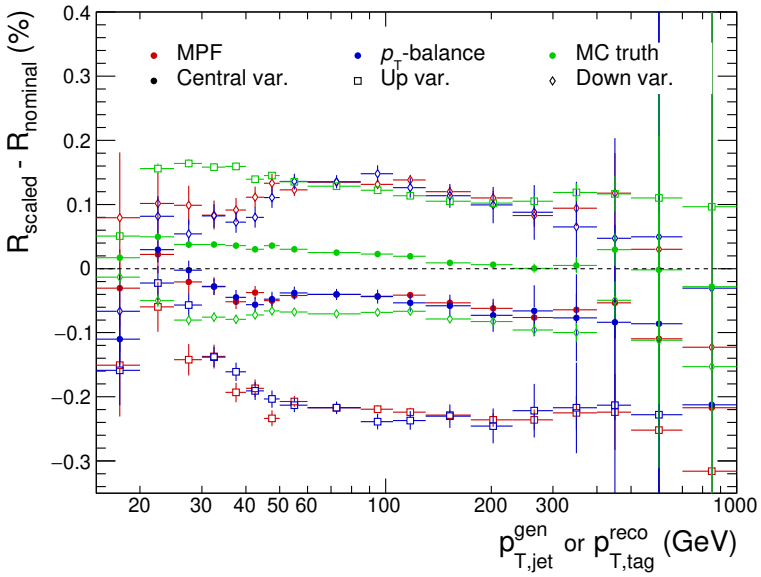

Figure 6.2: Change in the b jet responses, when the b-to-c hadron inclusive branching ratios were rescaled.

6.1.3 b hadron production fraction

The effect of b hadron production fraction scaling is shown in Figure 6.3. The MC truth response has shifted about -0.2% at $p_{\mathrm{T}} = 25$ GeV, after which the shift gets gradually smaller and reaches -0.1% level at $p_{\mathrm{T}} = 200$ GeV. For the MPF the effect is similar, as it starts with shift of -0.1% which disappears around $p_{\mathrm{T}} = 100$ GeV. The p_{T}-balance response changes accordingly but the shift is even smaller compared to MPF.

The difference between the MC truth and the experimental responses comes from the increased Λ_b^0 production. As mentioned before, the semileptonic decays of Λ_b^0 in P8 are more rare compared to the other b hadron types. Thus the production fraction rescaling actually decreases the neutrino production, which creates difference between the responses.

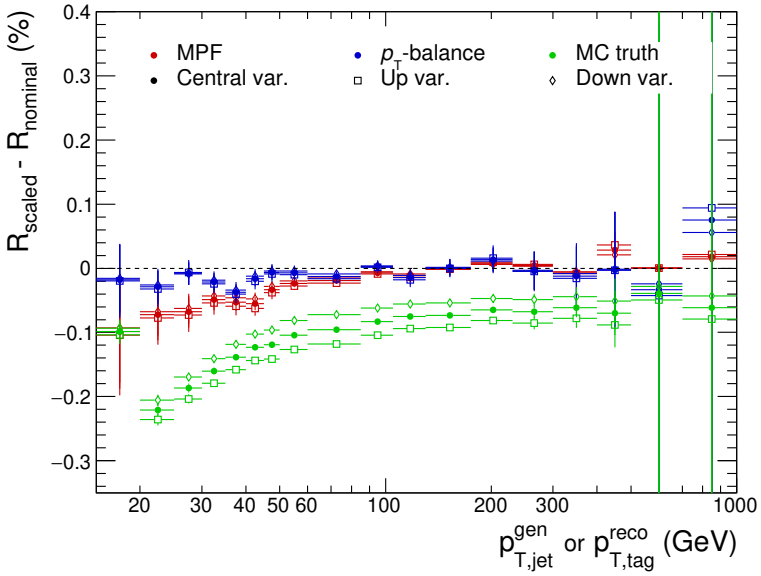

Figure 6.3: Change in the b jet responses, when the production fractions of different b hadrons were rescaled.

6.1.4 b quark fragmentation

The effect of scaling x_B distribution according to the newly parameterized b quark fragmentation function is shown in Figure 6.4. All the response changes are now in the range from -0.1 to +0.1%. Interestingly, all the three responses show a bit different behaviour compared to each other. The MC truth central variation shows very small positive shift, less than 0.02% at $p_{\mathrm{T}} < 30$ GeV, after which the response change is negligible. The MPF does not show any shift at $p_{\mathrm{T}} < 60$ GeV, but after this there is increasing positive shift, which reaches 0.05% level at $p_{\mathrm{T}} = 500$ GeV. The p_{T}-balance behaves otherwise similarly but starts from -0.05% level and increases to a bit higher level compared to MPF.

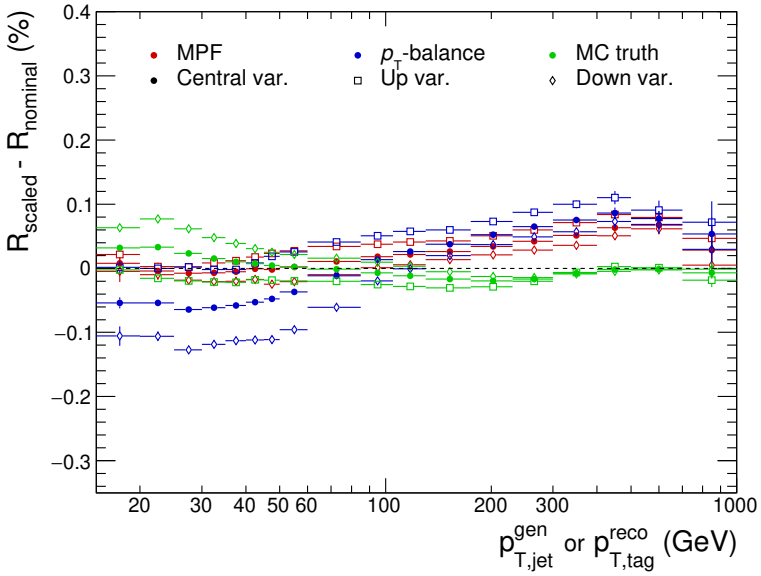

Figure 6.4: Change in the b jet responses, when the b quark fragmentation function was rescaled to new parameterization.

6.1.5 Summary plots of all the rescaling methods

Figures 6.5, 6.6 and 6.7 show the different scaling methods in one plot for MC truth, MPF and p_{T}-balance responses respectively. These plots contain also the total uncertainty envelope calculated from the total shift, as discussed in Section 5.4.5.

As seen from Figure 6.5, the b production fraction and the b-to-c hadron BR rescaling have the biggest effect on the MC truth response. The other scaling methods have only effect below 0.1% level. On the MPF and p_{T}-balance responses, shown in Figures 6.6 and 6.7, the b-to-c hadron BR rescaling produces the biggest shift. However, the b production fraction is now only significant at low p_{T} values and the b fragmentation scaling even exceeds it in the p_{T}-balance response.

The biggest up and down variations are produced by the b-to-c rescaling for all the response types. The b hadron semileptonic BR rescaling also contributes significantly to the up and down variations of the experimental responses.

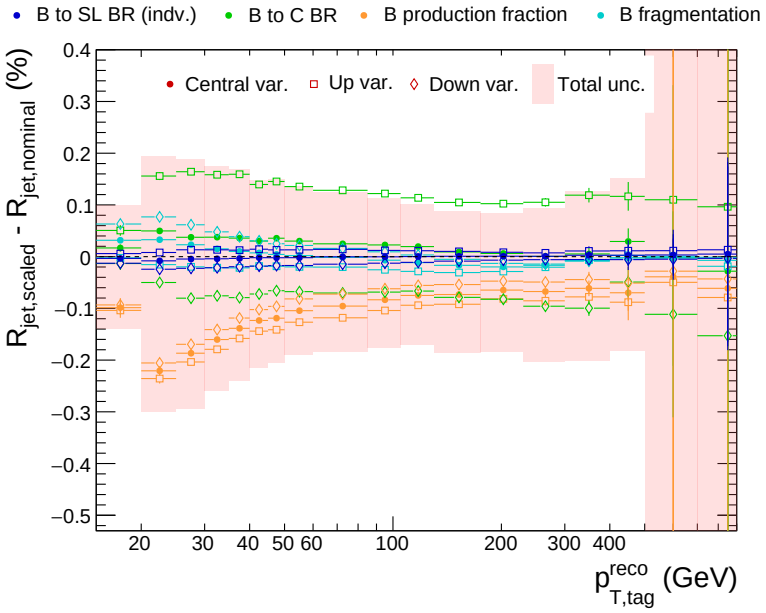

Figure 6.5: Change in the b jet MC truth response due to the different scaling methods applied individually. The total uncertainty band on the background is calculated from the sum of all the scaling methods.

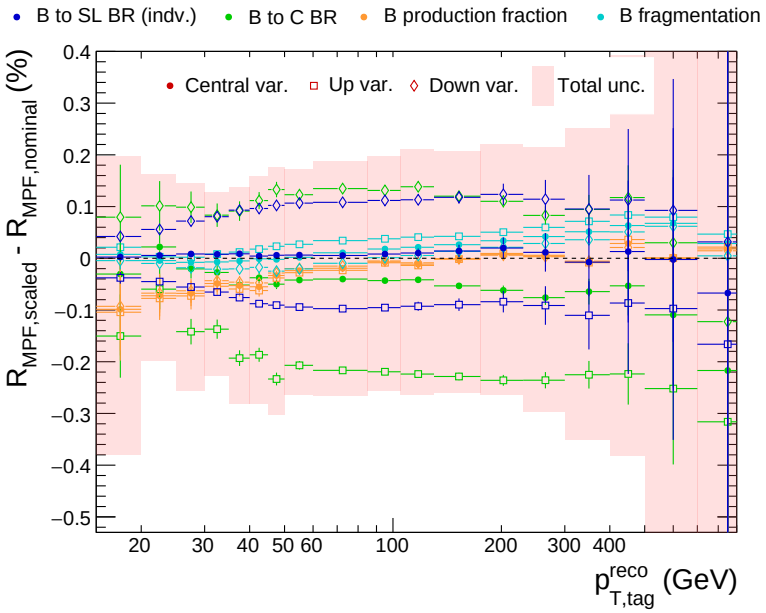

Figure 6.6: Change in the b jet MPF-response due to the different scaling methods applied individually. The total uncertainty band on the background is calculated from the sum of all the scaling methods.

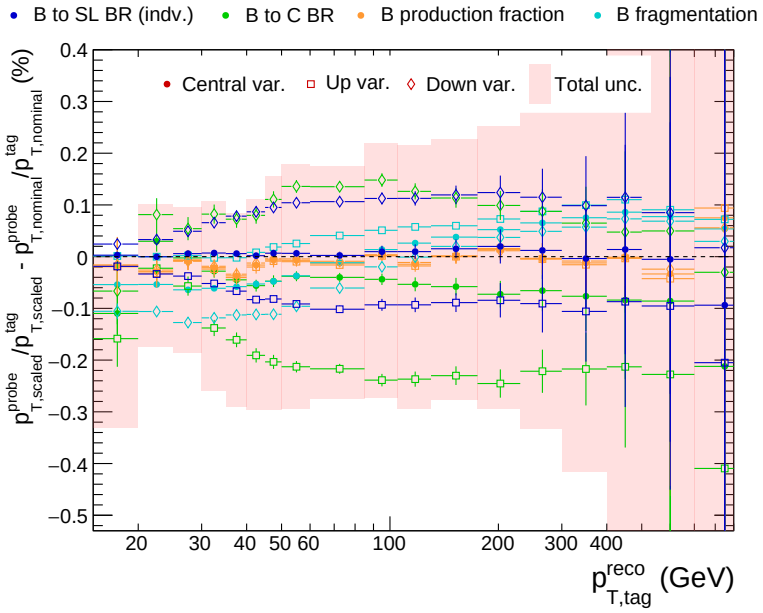

Figure 6.7: Change in the b jet p_{T}-balance response due to the different scaling methods applied individually. The total uncertainty band on the background is calculated from the sum of all the scaling methods.

6.2 Total effect on the jet response

The total effect on the jet responses when all the scaling methods are applied, is shown for MC truth, MPF and p_{T}-balance responses in Figures 6.8, 6.9 and 6.10, respectively. As seen from the plots, the central variation values predict a small shift downwards. This is most clearly seen on the MC truth response in Figure 6.8, where shift is on -0.15% level at $p_{\mathrm{T}} = 25$ GeV, after which is decreases to 0.1% at higher p_{T} values. The total uncertainties range from -0.3 to 0.2% at $p_{\mathrm{T}} = 25$ GeV and then decrease at higher p_{T}, and range from -0.2 to 0.1% at $p_{\mathrm{T}} = 300$ GeV.

The MPF and p_{T}-balance responses show -0.1% shift at $p_{\mathrm{T}} = 40$ GeV, which eventually vanishes around $p_{\mathrm{T}} = 150$ GeV when p_{T} increases. The shift also gets smaller as the p_{T} decreases from $p_{\mathrm{T}} = 40$ GeV, but does not vanish completely. The total uncertainties vary from -0.3 to 0.2% with the MPF. The total uncertainties of p_{T}-balance are similar although a bit smaller at $p_{\mathrm{T}} < 40$ GeV.

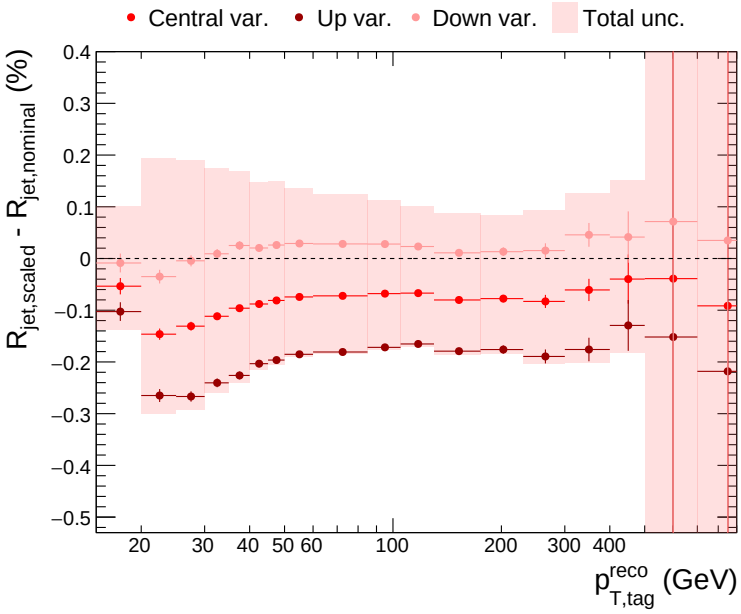

Figure 6.8: Total shift in the MC truth response when all the scaling methods are applied.

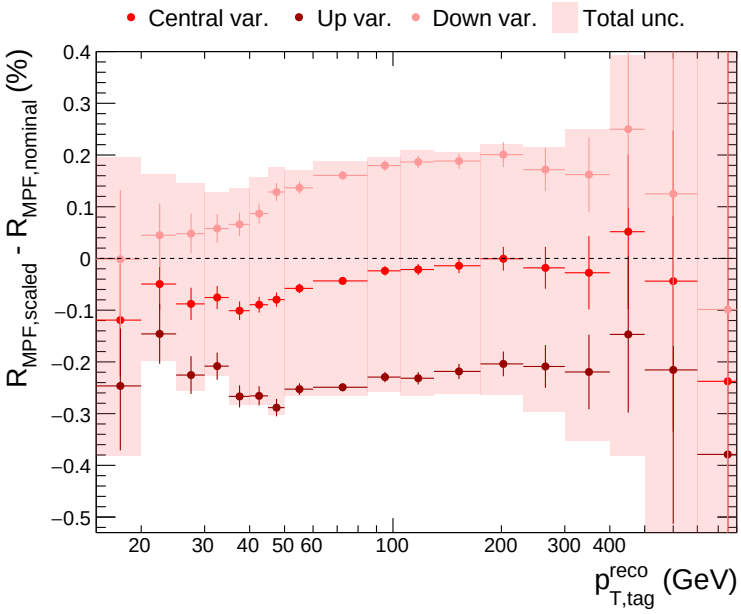

Figure 6.9: Total shift in the MPF-response when all the scaling methods are applied.

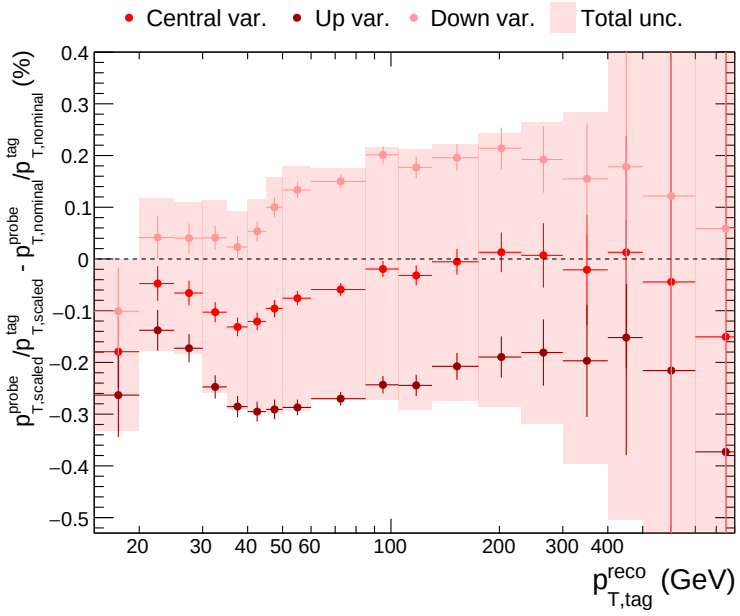

Figure 6.10: Total shift in the p_{T}-balance response when all the scaling methods are applied.

7. Discussion

In this chapter the effects of the different rescaling methods on JES are discussed one by one. Also several improvements are suggested to the analysis methods and some future research topics are considered. All the discussion focuses on the $p_{\mathrm{T}} < 300$ GeV range, where the precision is the best.

7.1 b hadron semileptonic branching ratio

The scaling of b hadron semileptonic branching ratios did not show considerable effect to the jet responses. The admixture scaling yielded small downshift to MPF and p_{T}-balance responses, as in the other scaling method where individual b hadron types were considered, there were practically no shift to be detected. The small shift with the admixture model is understandable as the associated scaling factors are rather small, as can be seen from Figure 5.3.

This not the case with the exclusive BR scaling, where the scaling factor of Λ_b^0 channel is approximately $c^{\Lambda_b^0} \approx 1.4$ for the whole p_{T} range. Now the non-existent change in the bJES can be explained with the rescaling of the $B^{\pm}$ channel, which scaling factor is on average $c^{B^{\pm}} \approx 0.93$, as can be calculated from Table 5.2. However, these balance each other out when we take the production fractions of P8 into account. The Λ_b^0 decays scales events up 40% but are produced only at rate of $f_{\Lambda_b^0} = 3.7\%$, while $B^{\pm}$ scales events 3% down and are produced at rate of $f_u = 42.9\%$, as can be seen from Table 5.4.

This hypothesis could be tested by rescaling the Λ_b^0 and $B^{\pm}$ decays separately and inspecting if they indeed produce equal shifts in opposite directions. As the production fraction of Λ_b^0 in P8 is approximately only half of the PDG reference value, the semileptonic Λ_b^0 decays should double their effect on bJES, if the production fraction rescaling is done beforehand. This is one possible subject for future studies.

7.2 b-to-c hadron inclusive branching ratios

The b-to-c hadron BR scaling seem to cause considerable shift to the bJES which is a unexpected result. The reason that b-to-c hadron BR scaling was considered in the first place was that the ATLAS study [63] scaled also the c hadron production fractions. However, as this thesis concentrated to the b jets, we did not want to complicate the analysis more by taking also the c-jets into account.

As large fraction of the produced c hadrons are coming from the b hadron decays, this allowed us to inspect and scale the c hadron production via the b jets. The large shift in the bJES, is naturally arising from the large scaling factors, which are due to the big differences between the reference values and P8. As can be seen from Figure 5.5, the differences are indeed very large, especially for the channel $B^+ \to D_s^+$. The possibility of some error in the BR calculations from the P8 cannot be excluded as the differences are so remarkable.

The results of b-to-c hadron BR scaling are also interesting. As seen from Figure 6.2, the MC truth response shifts up, as the MPF and p_{T}-balance responses go down. As the neutrinos are not included into MC truth response, it is natural to suspect that some phenomenon with neutrinos is causing the difference. This can be confirmed from Fig. G.3, where the MC truth response has shifted down towards the experimental responses after the generator level neutrinos were added to it.

The difference in the responses was studied further by excluding the rescaling of B^+ and B^0 decay modes separately and then calculating the responses again. These results are presented in Appendix C. As seen from Figure D.1, the small shift in the MC truth response is due to the B^+ modes. In the case of MPF and p_{T}-balance responses it seems that B^+ decays causes -0.1% shift as B^0 decays produce shift of +0.06%. The net shift of the normal scaling is thus dominated by the B^+ mode. The large systematic uncertainties in bJES can be pinpointed to the rather large uncertainties of the reference branching ratios. As can be seen from Table 5.3, especially the channels $B^0 \to D_s^+$, $B^+ \to D^+$ and $B^+ \to D_s^+$ have large relative uncertainties.

Generally, the response shift and large uncertainties may be caused by the large amount of neutral hadrons produced from the c hadron decays. Especially the D meson decays produce often neutral kaons, which can be reconstructed less precisely compared to the charged hadrons. This hypothesis is supported by Figure G.3 where we see that the MC truth response is not exactly on the same level with the experimental responses, although the neutrinos are added into it. This suggests that besides neutrinos some other mechanism is causing part of the shift.

One future improvement to the analysis would be including rescaling of the Λ_c^+ decays from the b hadrons. Actually the reference value of Λ_b^0 decay channel used in the exclusive b hadron semileptonic BR scaling is for $\mathrm{BR}(\Lambda_b^0 \to \Lambda_c^+ l^- \bar{\nu}_l + \text{anything})$ mode. Thus the effect of Λ_c^+ decays from b hadrons is taken into account partially already in the exclusive b hadron semileptonic BR scaling results.

7.3 b hadron production fractions

Scaling of the b hadron production fractions causes a negative shift to the MC truth response, however this shift is much smaller in MPF and p_{T}-balance responses. The shape of the shift, *i.e.* increasing at low p_{T}, suggests that the shift

is due to the scaling of production fraction of Λ_b^0 , which can be seen by comparing Figures 5.8 and 6.3. The reason between the responses can be pinpointed again to the neutrino production. This can be seen from Figure G.4 in which the MC truth response almost coincides with the MPF when the neutrinos are added to it. As Λ_b^0 decays semileptonically less compared to other b hadrons, the increased Λ_b^0 production actually decreases the semileptonic decays which creates difference between the responses. This hypothesis could be tested by conducting the b hadron semileptonic BR rescaling simulataneously with the production fraction rescaling and observing whether the difference between MC truth and experimental responses is still present.

It is also good to keep in mind that the study made by LHCb [66] is one of the first detecting p_{T} -dependent production fractions of b hadrons. Further studies are also required to confirm these findings. The LHCb measurement was conducted for generic pp collisions and only for low p_{T} values. We used only Z+jet events and extrapolated the results to higher p_{T} -values, which might bias the results.

We found out that the fit function for B_s^0 used by the LHCb is not suitable for the higher p_{T}-values. Also reference values used by ATLAS for their top mass measurement [63], were found to be questionable. The values used in that paper were global averages, which were discontinued in the latest HVFLAV publication [28], as the production fractions are dependent on the production environment. Thus the usage of fractions measured at the LHC were favoured in this thesis. It should be also noted that the p_{T} dependency of the production fractions might differ between different event types. Thus, before applying these results to $t\bar{t}$ events, the possible p_{T} dependency of production fractions should be evaluated for this event type.

7.4 b quark fragmentation

The effect of the fragmentation scaling on bJES seems to be below 0.1%. Only a small shift was expected as the reference value for r_B was very close to the default value used by the P8. As seen in Figure 5.9, the biggest scaling factors are applied to the small and very high x_B-values, which are more rare, compared to the upper mid-range where most of the events are located. This also explains the smallness of the bJES shift.

There were some complications in the production of the x_B distribution. In order to get the distribution mostly under the 1.0 value, the neutrinos inside the jet and the particles located outside of the jet cone, decayed from the b hadron, were added to the generator level jet p_{T}. However, still part of the tail of the x_B distribution was over the 1.0 value, which correspond to situation where b hadron inside the jet has larger p_{T} compared to the jet.

Also when compared to the x_B distribution produced in the CMS study [67], from which the reference value for r_B parameter was taken, we did not manage to reproduce similar distribution shape. This can be seen in Figure C.3.

There are multiple reasons that can explain the differences between the distributions. First and foremost, the CMS study used $t\bar{t}$ single and di-lepton channels for the measurement. Compared to Z+jet events, $t\bar{t}$ events contain bigger background from UE and also the b jets produced in the $t\bar{t}$ have usually higher momentum. In addition, the CMS study used slightly different p_{T} and η cuts, $p_{\mathrm{T}} > 30$ GeV and $\eta < 2.4$, compared to the $p_{\mathrm{T}} > 15$ GeV and $\eta < 1.3$ used in this thesis. As shown in Fig. C.4, the distribution moves to the left, closer to the reference, as the p_{T} increases. Thus the higher jet p_{T} associated with the $t\bar{t}$ events and higher jet p_{T} cut, could be one explanation for the shape difference of the distributions.

It seems to be common to add the neutrinos to the generator level jets in the b fragmentation studies, as this was done also in [65] and [67]. But it was not clear whether the b hadron decay products outside of the b jet cone were also considered in these cases. The interpretation of the b hadron p_{T} differed at least. In our study the b hadron p_{T} was taken from the generator level hadron, in contrast to [67], where the p_{T} was calculated from its decay products. The neutrino p_{T} is not taken into account in the latter approach, which causes the x_B values to be smaller. This might explain why the reference x_B distribution was flatter and shifted to the left compared to ours.

Another reason explaining the differences is related to the tunes used in the studies. As noted earlier, the CMS study used P8 tune `CUETP8M1`, while in this study we did `CP5`. This might also bias the scaling. Thus additional studies need to be made to confirm the results related to the fragmentation function scaling. One option would be redoing the scaling with same event type and cuts used in the CMS study. Also the effect of reconstruction of the b hadron p_{T} from its decay products and the usage of different tune should investigated more thoroughly.

7.5 Further improvements to the methods and future research topics

In order to cross-check the results, it would be useful to generate the data with Herwig 7 and do the rescaling again. If similar JES shifts would be obtained also with the H7, it would be a solid proof that there is actually some physical phenomenon or common MC simulator feature behind the shift.

The error estimation of the analysis should also be revised, as the systematic errors related to the production fraction and fragmentation scaling are suspiciously small. In addition, the systematic errors in general, interpreted as simultaneous variation of all the scaling parameters up and down by one standard deviation, could be studied more carefully. It is possible that larger errors are produced by some combination of up and down variations, so all the possible combinations should be evaluated.

All the different rescaling methods used in this thesis could be also tested for different hard process event types. Especially, it should be investigated using

$t\bar{t}$-events, if these results were applied on jets to the analysis of top quark mass measurements. This would also help to understand if the shifts seen on the JES are characteristic only for the Z+b events or common for all event types containing b jets.

All in all, the rescaling method presented in this thesis could be utilized also for studies of other quark flavours. The features used in the rescaling should be then considered again. Especially interesting research subject would be the gluon jets as they cause largest uncertainties to JES calculations at the CMS for the time being.

8. Conclusions

The combined effect of all four different rescaling features suggests that the bJES is known at 0.2% level precision at the p_{T} range below 300 GeV. Also a small shift of -0.1% in the bJES is predicted for MPF and p_{T}-balance responses at low p_{T} range, after which the effect gradually vanishes with the increasing p_{T}.

The main contributions to the uncertainty of the bJES come from rescaling the branching ratios of semileptonic b hadron and b-to-c hadron decays. The effect of neutrino production in semileptonic b decays was confirmed by clustering neutrinos also to the MC truth response, which shifted it in a way that reproduced the MPF and p_{T}-balance shifts. The effect of b-to-c hadron BR rescaling was also confirmed to be partly due to the neutrino production, and partly due to the production of neutral hadrons via D meson decays, although this was not fully confirmed.

The small shift in the central variation of experimental responses was mostly created by rescaling of the b hadron production fractions and b-to-c hadron branching ratios. Especially rescaling the production fraction of Λ_b^0 baryons played an important role, as the experimental reference values suggested that the production should be almost twice as large as it is in P8 by default. Also the semileptonic BR of Λ_b^0 should be scaled up 40% in P8 according to references. Thus both of these properties should be rescaled in P8 in order to fix the imprecise Λ_b^0 modeling.

There are many unclear aspects related to the rescaling methods, as pointed out in the previous chapter. The exact physical mechanisms by which the different rescaling methods affect to the bJES should be studied in greater detail, especially in the case of b-to-c hadron BR rescaling as it had such a big effect on the bJES. Nevertheless, this thesis provided an overview to the different features affecting to the bJES and tried to evaluate their effect on the energy scale with high precision.

A possible 0.2% level precision of bJES suggested in this thesis, improves remarkably 0.4-0.5% JES accuracy achieved at CMS during Run 1 of the LHC [3]. This high level of precision in the jet calibration is especially needed, when High Luminosity LHC (HL-LHC) begins its operation. As the HL-LHC will produce about 20 times as much data as has been produced by now at the LHC, the statistical precision of many analyses increase significantly. If the 0.1% precision of JES is not achieved, the jet calibration could create a bottleneck in the follow-up studies.

If the 0.2% level bJES accuracy is confirmed in future studies, the analysis

moving towards the precision measurement with more data, sensitive to b jet physics, *e.g.* top quark mass measurement, will take a huge leap forward and hopefully yield many interesting results and discoveries.

Bibliography

[1] CMS collaboration. "Observation of a New Boson at a Mass of 125 GeV with the CMS Experiment at the LHC". In: *Phys. Lett. B* 716 (2012), pp. 30–61. DOI: 10.1016/j.physletb.2012.08.021. arXiv: 1207.7235 [hep-ex].

[2] ATLAS collaboration. "Observation of a new particle in the search for the Standard Model Higgs boson with the ATLAS detector at the LHC". In: *Physics Letters B* 716.1 (Sept. 2012), pp. 1–29. ISSN: 0370-2693. DOI: 10.1016/j.physletb.2012.08.020.

[3] CMS collaboration. "Jet energy scale and resolution in the CMS experiment in pp collisions at 8 TeV". In: *Journal of Instrumentation* 12.02 (Feb. 2017), P02014–P02014. ISSN: 1748-0221. DOI: 10.1088/1748-0221/12/02/p02014.

[4] S. Alekhin, A. Djouadi, and S. Moch. "The top quark and Higgs boson masses and the stability of the electroweak vacuum". In: *Physics Letters B* 716.1 (Sept. 2012), pp. 214–219. ISSN: 0370-2693. DOI: 10.1016/j.physletb.2012.08.024.

[5] T. Sjöstrand et al. "An introduction to PYTHIA 8.2". In: *Computer Physics Communications* 191 (June 2015), pp. 159–177. ISSN: 0010-4655. DOI: 10.1016/j.cpc.2015.01.024.

[6] CMS collaboration. "The CMS Experiment at the CERN LHC". In: *JINST* 3 (2008), S08004. DOI: 10.1088/1748-0221/3/08/S08004.

[7] T. Han. "Collider Phenomenology: Basic Knowledge and Techniques". In: *Physics In D 4 Tasi 2004* (July 2006). DOI: 10.1142/9789812773579_0008.

[8] M. Thomson. *Modern particle physics*. Physics textbook. New York: Cambridge University Press, 2013. ISBN: 978-1-107-03426-6.

[9] D. Griffiths. *Introduction to elementary particles; 2nd rev. version*. Physics textbook. New York, NY: Wiley, 2008. ISBN: 978-3-527-40601-2.

[10] Lecture note of the course "Lectures on Particle Physics". http://www.damtp.cam.ac.uk/user/tong/particle.html. Accessed: 2021-08-01.

[11] P. A. Zyla et al. "Review of Particle Physics". In: *PTEP* 2020.8 (2020 and 2021 update), p. 083C01. DOI: 10.1093/ptep/ptaa104.

[12] M. Peskin and D. Schroeder. *An introduction to quantum field theory*. Boulder, CO: Westview, 1995. ISBN: 978-0-201-50397-5.

[13] Wikipedia entry for the Standard Model. https://en.wikipedia.org/wiki/Standard_Model. Accessed: 2021-08-01.

[14] G. Degrassi et al. "Higgs mass and vacuum stability in the Standard Model at NNLO". In: *Journal of High Energy Physics* 2012.8 (Aug. 2012). ISSN: 1029-8479. DOI: 10.1007/jhep08(2012)098.

[15] S. Alekhin, A. Djouadi, and S. Moch. "The top quark and Higgs boson masses and the stability of the electroweak vacuum". In: *Physics Letters B* 716.1 (Sept. 2012), pp. 214–219. ISSN: 0370-2693. DOI: 10.1016/j.physletb.2012.08.024.

[16] A. Andreassen, W. Frost, and M. D. Schwartz. "Scale-invariant instantons and the complete lifetime of the Standard Model". In: *Physical Review D* 97.5 (Mar. 2018). ISSN: 2470-0029. DOI: 10.1103/physrevd.97.056006.

[17] J. Bellm et al. "Herwig 7.0/Herwig++ 3.0 release note". In: *Eur. Phys. J. C* 76.4 (2016), p. 196. DOI: 10.1140/epjc/s10052-016-4018-8. arXiv: 1512.01178 [hep-ph].

[18] L. N. Lipatov. "The parton model and perturbation theory". In: *Yad. Fiz.* 20 (1974), pp. 181–198.

[19] V. N. Gribov and L. N. Lipatov. "Deep inelastic ep scattering in perturbation theory". In: *Sov. J. Nucl. Phys.* 15 (1972), pp. 438–450.

[20] G. Altarelli and G. Parisi. "Asymptotic Freedom in Parton Language". In: *Nucl. Phys. B* 126 (1977), pp. 298–318. DOI: 10.1016/0550-3213(77)90384-4.

[21] Y. Dokshitzer. "Calculation of the Structure Functions for Deep Inelastic Scattering and e+e- Annihilation by Perturbation Theory in Quantum Chromodynamics." In: *Sov. Phys. JETP* 46 (1977), pp. 641–653.

[22] R. D. Ball et al. "Parton distributions from high-precision collider data". In: *The European Physical Journal C* 77.10 (Oct. 2017). ISSN: 1434-6052. DOI: 10.1140/epjc/s10052-017-5199-5.

[23] A. Buckley et al. "General-purpose event generators for LHC physics". In: *Physics Reports* 504.5 (July 2011), pp. 145–233. ISSN: 0370-1573. DOI: 10.1016/j.physrep.2011.03.005.

[24] T. Sjöstrand, S. Mrenna, and P. Skands. "PYTHIA 6.4 physics and manual". In: *Journal of High Energy Physics* 2006.05 (May 2006), pp. 026–026. ISSN: 1029-8479. DOI: 10.1088/1126-6708/2006/05/026.

[25] B. Andersson et al. "Parton fragmentation and string dynamics". In: *Physics Reports* 97.2 (1983), pp. 31–145. ISSN: 0370-1573. DOI: https://doi.org/10.1016/0370-1573(83)90080-7.

[26] B. Andersson. *The Lund model*. Vol. 7. Cambridge University Press, July 2005. ISBN: 978-0-521-01734-3.

[27] M. G. Bowler. "e+ e- Production of Heavy Quarks in the String Model". In: *Z. Phys. C* 11 (1981), p. 169. DOI: 10.1007/BF01574001.

[28] Y. Amhis et al. "Averages of b-hadron, c-hadron, and τ-lepton properties as of 2018". In: *The European Physical Journal C* 81.3 (Mar. 2021). ISSN: 1434-6052. DOI: 10.1140/epjc/s10052-020-8156-7.

[29] J. Beringer et al. "Review of Particle Physics (RPP)". In: *Phys. Rev. D* 86 (2012), p. 010001. DOI: 10.1103/PhysRevD.86.010001.

[30] CMS collaboration. "Extraction and validation of a new set of CMS Pythia8 tunes from underlying-event measurements". In: *The European Physical Journal C* 80.1 (Jan. 2020). ISSN: 1434-6052. DOI: 10.1140/epjc/s10052-019-7499-4.

[31] O. S. Brüning et al. *LHC Design Report*. CERN Yellow Reports: Monographs. Geneva: CERN, 2004. DOI: 10.5170/CERN-2004-003-V-1. URL: https://cds.cern.ch/record/782076.

[32] L. Evans. "The Large Hadron Collider". In: *Ann. Rev. Nucl. Part. Sci.* 61 (2011), pp. 435–466. DOI: 10.1146/annurev-nucl-102010-130438.

[33] J. Gruschke. "Observation of Top Quarks and First Measurement of the $t\bar{t}$- Production Cross Section at a Centre-Of-Mass Energy of 7 TeV with the CMS Experiment at the LHC". PhD thesis. Karlsruhe U., EKP, 2011. DOI: 10.5445/IR/1000022394.

[34] J. Jowett. "Colliding Heavy Ions in the LHC". In: *9th International Particle Accelerator Conference*. June 2018. DOI: 10.18429/JACoW-IPAC2018-TUXGBD2.

[35] CMS TWiki describing the luminosity recorded by the CMS and produced the LHC. https://twiki.cern.ch/twiki/bin/view/CMSPublic/LumiPublicResults#Multi_year_plots. Accessed: 2021-05-20.

[36] Website presenting long term LHC Run schedule. https://lhc-commissioning.web.cern.ch/schedule/LHC-long-term.htm. Accessed: 2021-05-19.

[37] G. Apollinari et al. "High-Luminosity Large Hadron Collider (HL-LHC)". Technical Design Report V. 0.1, CERN-2017-007-M. Apr. 2017. DOI: 10.23731/CYRM-2017-004.

[38] Official website of the CERN reporting about all the new hadrons found during the first 10 operational years of the LHC. https://home.cern/news/news/physics/59-new-hadrons-and-counting. Accessed: 2021-05-20.

[39] CMS collaboration. "The CMS experiment at the CERN LHC". In: *JINST* 3 (2008), S08004. DOI: 10.1088/1748-0221/3/08/S08004.

[40] CMS collaboration. "Particle-flow reconstruction and global event description with the CMS detector". In: *JINST* 12 (2017), P10003. DOI: 10.1088/1748-0221/12/10/P10003. arXiv: 1706.04965 [physics.ins-det].

[41] Section in the official CMS webpage describing how the detector works. https://cms.cern/detector. Accessed: 2021-05-21.

[42] CMS collaboration. "Performance of the CMS Level-1 trigger in proton-proton collisions at $\sqrt{s} = 13\,\text{TeV}$". In: *JINST* 15 (2020), P10017. DOI: 10.1088/1748-0221/15/10/P10017. arXiv: 2006.10165 [hep-ex].

[43] CMS collaboration. "The CMS trigger system". In: *JINST* 12 (2017), P01020. DOI: 10.1088/1748-0221/12/01/P01020. arXiv: 1609.02366 [physics.ins-det].

[44] CMS collaboration. "Performance of missing transverse momentum reconstruction in proton-proton collisions at $\sqrt{s} = 13\,\text{TeV}$ using the CMS detector". In: *JINST* 14 (2019), P07004. DOI: 10.1088/1748-0221/14/07/P07004. arXiv: 1903.06078 [hep-ex].

[45] S. Catani and D. Zeppenfeld. "Jet Algorithms". In: *1st Les Houches Workshop on Physics at TeV Colliders*. May 2000.

[46] M. Cacciari et al. "FastJet user manual". In: *Eur. Phys. J. C* 72 (2012), p. 1896. DOI: 10.1140/epjc/s10052-012-1896-2. arXiv: 1111.6097 [hep-ph].

[47] Y.Dokshitzer et al. "Better jet clustering algorithms". In: *Journal of High Energy Physics* 1997.08 (Aug. 1997), pp. 001–001. ISSN: 1029-8479. DOI: 10.1088/1126-6708/1997/08/001.

[48] S. Ellis and E. Davison. "Successive combination jet algorithm for hadron collisions". In: *Physical Review D* 48.7 (Oct. 1993), pp. 3160–3166. ISSN: 0556-2821. DOI: 10.1103/physrevd.48.3160.

[49] M. Cacciari, G. Salam, and G. Soyez. "The anti-k_t jet clustering algorithm". In: *JHEP* 04 (2008), p. 063. DOI: 10.1088/1126-6708/2008/04/063. arXiv: 0802.1189 [hep-ex].

[50] CMS collaboration. "Determination of jet energy calibration and transverse momentum resolution in CMS". In: *Journal of Instrumentation* 6.11 (Nov. 2011), P11002–P11002. ISSN: 1748-0221. DOI: 10.1088/1748-0221/6/11/p11002.

[51] ATLAS collaboration. "Jet energy measurement with the ATLAS detector in proton-proton collisions at $\sqrt{\text{s}} = 7$ TeV". In: *The European Physical Journal C* 73.3 (Mar. 2013). ISSN: 1434-6052. DOI: 10.1140/epjc/s10052-013-2304-2.

[52] CMS collaboration. "Jet energy scale and resolution performance with 13 TeV data collected by CMS in 2016-2018". CMS analysis note, CMS-DP-2020/019. Apr. 2020. URL: https://cds.cern.ch/record/2715872.

[53] CMS collaboration. "Identification of heavy-flavour jets with the CMS detector in pp collisions at 13 TeV". In: *Journal of Instrumentation* 13.05 (May 2018), P05011–P05011. ISSN: 1748-0221. DOI: 10.1088/1748-0221/13/05/p05011.

[54] S. Agostinelli et al. "GEANT4–a simulation toolkit". In: *Nucl. Instrum. Meth. A* 506 (2003), pp. 250–303. DOI: 10.1016/S0168-9002(03)01368-8.

[55] S. Banerjee. "CMS simulation software". In: *J. Phys. Conf. Ser.* 396 (2012). Ed. by Michael Ernst et al., p. 022003. DOI: 10.1088/1742-6596/396/2/022003.

[56] R. Rahmat, R. Kroeger, and A. Giammanco. "The fast simulation of the CMS experiment". In: *J. Phys. Conf. Ser.* 396 (2012). Ed. by Michael Ernst et al., p. 062016. DOI: 10.1088/1742-6596/396/6/062016.

[57] J. Favereau et al. "DELPHES 3: a modular framework for fast simulation of a generic collider experiment". In: *Journal of High Energy Physics* 2014.2 (Feb. 2014). ISSN: 1029-8479. DOI: 10.1007/jhep02(2014)057.

[58] M. Myllymäki. "Parameterized model of Particle Flow event reconstruction for jet energy scale at CMS experiment". English. Master's thesis. Aalto University. School of Science, 2020, pp. 64 + 14. URL: http://urn.fi/URN:NBN:fi:aalto-202008234990.

[59] A. Buckley et al. "LHAPDF6: parton density access in the LHC precision era". In: *The European Physical Journal C* 75.3 (Mar. 2015). ISSN: 1434-6052. DOI: 10.1140/epjc/s10052-015-3318-8.

[60] R. Brun and F. Rademakers. "ROOT: An object oriented data analysis framework". In: *Nucl. Instrum. Meth. A* 389 (1997). Ed. by M. Werlen and D. Perret-Gallix, pp. 81–86. DOI: 10.1016/S0168-9002(97)00048-X.

[61] D0 collaboration. "Jet Energy Scale Determination in the D0 Experiment". In: *Nucl. Instrum. Meth. A* 763 (2014), pp. 442–475. DOI: 10.1016/j.nima.2014.05.044. arXiv: 1312.6873 [hep-ex].

[62] T. Mäkelä. "Flavour-dependent jet energy corrections and top quark mass". English. Master's thesis. Aalto University. School of Science, 2019, pp. 53 + 40. URL: http://urn.fi/URN:NBN:fi:aalto-201906234090.

[63] ATLAS Collaboration. "Measurement of the top quark mass using a leptonic invariant mass in pp collisions at $\sqrt{s} = 13$ TeV with the ATLAS detector". ATLAS CONF Note, ATLAS-CONF-2019-046. Oct. 2019. URL: https://cds.cern.ch/record/2693954.

[64] M. Tanabashi et al. "Review of Particle Physics". In: *Phys. Rev. D* 98 (3 Aug. 2018), p. 030001. DOI: 10.1103/PhysRevD.98.030001.

[65] CMS official guidelines for evaluating the main sources of systematic uncertainties in t-quark analyses in Run 2. https://twiki.cern.ch/twiki/bin/viewauth/CMS/TopSystematics. Accessed: 2021-06-04.

[66] LHCb collaboration. "Measurement of b hadron fractions in 13 TeV pp collisions". In: *Phys. Rev. D* 100.3 (2019), p. 031102. DOI: 10.1103/PhysRevD.100.031102. arXiv: 1902.06794 [hep-ex].

[67] CMS Collaboration. "Measurement of the shape of the b quark fragmentation function using charmed mesons produced inside b jets from $t\bar{t}$ pair decays". CMS Physics Analysis Summary, CMS-PAS-TOP-18-012. Aug. 2021. URL: https://cds.cern.ch/record/2771694.

A. Scaling factor calculations

In this section, same notation is used for all the rescaling methods. Let us denote the reference fraction as f_x, where the x is particular decay channel or production fraction in the question. As subscript, the x will serve same purpose for other variables as well. The sum of individual events, or more precisely the sum of event weights, in some particular category, is then denoted as n_x.

The scaling factors are denoted as c_x. The sum of event weights before the scaling is defined as $\sum_x n_x = N^i_{tot}$ and after the scaling, $\sum_x c_x n_x = N^f_{tot}$. It is worth to mention that in the branching ratio studies, it is possible that the b hadron do not decay into any of the categories of interest. These cases form their own category *non-sl*, and they also contribute to the total event weights N^i_{tot} and N^f_{tot}. As the scaling factors are calculated separately for each p_{T} bin, all the calculations in this section demonstrates the calculation procedure for one individual p_{T} bin.

A.1 b hadron semileptonic branching ratios

Let us consider first the case with mixture of different b hadron types. There are four different categories in which semileptonically decaying b hadron can belong to. These four categories correspond to decay chains: $b \to \mu/e$, $b \to \tau \to \mu/e$, $b \to c \to \mu/e$ and $b \to \bar{c} \to \mu/e$. We can define shorthand for them as l, τl, cl and $\bar{c}l$ respectively.

It is also possible that one b hadron decay contributes to two categories, *e.g.* first decaying semileptonicically to c hadron, which will subsequently decay semileptonically. Now if we demand that after the scaling, the branching ratio calculated from the MC sample is equal to reference value for each category, and that sum of the event weights do not change after the scaling, we get following non-linear system of equations

$$\begin{cases} f_l - \frac{c_l n_l + c_l c_{cl} n_{l,cl} + c_l c_{\bar{c}l} n_{l,\bar{c}l}}{N^f_{tot}} = 0 \\ f_{\tau l} - \frac{c_{\tau l} n_{\tau l} + c_{\tau l} c_{cl} n_{\tau l,cl} + c_{\tau l} c_{\bar{c}l} n_{\tau l,\bar{c}l}}{N^f_{tot}} = 0 \\ f_{cl} - \frac{c_{cl} n_{cl} + c_l c_{cl} n_{l,cl} + c_{\tau l} c_{cl} n_{\tau l,cl} + c_{cl} c_{\bar{c}l} n_{cl,\bar{c}l}}{N^f_{tot}} = 0 \\ f_{\bar{c}l} - \frac{c_{\bar{c}l} n_{\bar{c}l} + c_l c_{\bar{c}l} n_{l,\bar{c}l} + c_{\tau l} c_{\bar{c}l} n_{\tau l,\bar{c}l} + c_{cl} c_{\bar{c}l} n_{cl,\bar{c}l}}{N^f_{tot}} = 0 \\ N^i_{tot} - N^f_{tot} = 0, \end{cases} \tag{A.1}$$

where sum of initial weights is

$$N^i_{tot} = n_l + n_{\tau l} + n_{cl} + n_{\bar{c}l} + n_{l,cl} + n_{l,\bar{c}l} + n_{\tau l,cl} + n_{\tau l,\bar{c}l} + n_{cl,\bar{c}l} + n_{\text{non-sl}}, \quad \text{(A.2)}$$

and after the scaling

$$\begin{aligned} N^f_{tot} = c_l n_l + c_{\tau l} n_{\tau l} + c_{cl} n_{cl} + c_{\bar{c}l} n_{\bar{c}l} + c_l c_{cl} n_{l,cl} + c_l c_{\bar{c}} n_{l,\bar{c}l} + \\ c_{\tau l} c_{cl} n_{\tau l,cl} + c_{\tau l} c_{\bar{c}l} n_{\tau l,\bar{c}l} + c_{cl} c_{\bar{c}l} n_{cl,\bar{c}l} + c_{\text{non-sl}} n_{\text{non-sl}}. \end{aligned} \quad \text{(A.3)}$$

The scaling factors c_x can be then solved numerically from equation A.1 using ROOT `multiRootFinder` function.

Let us consider next he other semileptonic b hadron branching ratio scaling, where we are interested in decays of different b hadron types. The studied b hadrons are $B^{\pm}$, B^0 and B^0_s mesons and Λ^0_b baryons.

This time the scaling factors can be solved analytically. For each of the b hadron type, it is demanded that the total amount of particles of that type do not change. This will lead to two scaling factors for each category: one scaling the semileptonic decays and one scaling all the other decays. The scaling factors can be solved from the following system of equations

$$\begin{cases} f_x = \frac{c_x n_x}{N^i_{tot}} \\ 1 - f_x = \frac{c_{x,\text{non-sl}} n_{x,\text{non-sl}}}{N^i_{tot}}, \end{cases} \quad \text{(A.4)}$$

from which the scale factor can be solved easily for the semileptonic cases

$$c_x = \frac{N^i_{tot} f_x}{n_x}, \quad \text{(A.5)}$$

and for the non-semileptonic cases

$$c_{x,\text{non-sl}} = \frac{N^i_{tot}(1 - f_x)}{n_{x,\text{non-sl}}}, \quad \text{(A.6)}$$

where the total sum of weights are $N^i_{tot} = n_x + n_{x,\text{non-sl}}$ and $N^i_{tot} = c_x n_x + c_{x,\text{non-sl}} n_{x,\text{non-sl}}$.

A.2 b-to-c hadron inclusive branching ratios

Calculating the scaling factors for B^0 and B^+ to c hadron branching ratios happens in the same fashion as the b hadron admixture calculation in section A.1. Let us consider first the decay paths $B^+ \to D^+$, $B^+ \to D^0$ and $B^+ \to D^+_s$. We can define abbreviations for them as $+$, 0 and s respectively.

Again, we get system of non-linear equations

$$\begin{cases} f_+ - \frac{c_+ n_+ + c_+ c_0 n_{+,0} + c_+ c_s n_{+,s}}{N^i_{tot}} = 0 \\ f_0 - \frac{c_0 n_0 + c_+ c_0 n_{+,0} + c_+ c_{0s} n_{0,s}}{N^i_{tot}} = 0 \\ f_s - \frac{c_s n_s + c_+ c_s n_{+,s} + c_+ c_{0s} n_{0,s}}{N^i_{tot}} = 0 \\ N^i_{tot} - N^f_{tot} = 0, \end{cases} \quad \text{(A.7)}$$

where total sum of weights before the scaling is

$$N^i_{tot} = n_+ + n_0 + n_s + n_{+,0} + n_{+,s} + n_{0,s} + n_{\text{other}}, \tag{A.8}$$

and after the scaling

$$\begin{aligned} N^i_{tot} = {} & c_+ n_+ + c_0 n_0 + c_s n_s + c_+ c_0 n_{+,0} + \\ & c_+ c_s n_{+,s} + c_0 c_s n_{0,s} + c_{\text{other}} n_{\text{other}}. \end{aligned} \tag{A.9}$$

Again the scaling factors c_x can be solved numerically using ROOT `multiRootFinder` function. Exactly similar system of equations, as equation A.7, could be written for B^0.

The scaling factors for B^0_s and Λ^0_b can be calculated analytically by using the equations A.5 and A.6.

A.3 b hadron production fraction

Let the g_{0s} and g_{Λ_b} be the fits made to LHCb data for the normalized production fractions of B^0_s and Λ^0_b ($\frac{f_s}{f_u+f_d}$ and $\frac{f_{\Lambda_b}}{f_u+f_d}$). The fit values are

$$g_{0s} = 0.18628 \tag{A.10}$$

$$g_{\Lambda_b} = 8.37462 \cdot 10^{-2} + \exp(-0.971313 - 0.115054 \cdot x) \tag{A.11}$$

Let us denote the production fractions of B_0, B_+, B^0_s and Λ^0_b as f_+, f_0, f_{0s} and f_{Λ_b} respectively. All the other B mesons expect the B_0, B_+, B^0_s are marked as f_m, and all the other b baryons expect Λ^0_b as f_b.

We demand that production fractions sum to unity after the scaling, and that common scaling factor is used for the pairs: f_+ and f_0, f_{0s} and f_m, f_{Λ_b} and f_b. Production fractions must also change according to g_{0s} and g_{Λ_b}. These conditions can be written as following system of equations

$$\begin{cases} c_{0,+}(f_0 + f_+) + c_{0s,m}(f_{0s} + f_m) + c_{\Lambda_b,b}(f_{\Lambda_b} + f_b) - 1 = 0 \\ g_{0s} - \frac{c_{0s,m} f_{0s}}{c_{0,+}(f_0+f_+)} = 0 \\ g_{\Lambda_b} - \frac{c_{\Lambda_b,b} f_{0s}}{c_{0,+}(f_0+f_+)} = 0, \end{cases} \tag{A.12}$$

from which we can solve analytically the scaling factors.

For B^0 and B^+ the scaling factor is

$$c_{0+} = \frac{n_{\Lambda_b} n_{0s} n_{tot}}{(n_0 + n_+)(g_{0s} n_{\Lambda_b} n_m + g_{0s} n_{\Lambda_b} n_{0s} + g_{\Lambda_b} n_{0s} n_b + g_{\Lambda_b} n_{\Lambda_b} n_{0s} + n_{\Lambda_b} n_{0s})} \tag{A.13}$$

For simplicity, let us denote

$$\alpha = g_{0s} n_{\Lambda_b} n_m + g_{0s} n_{\Lambda_b} n_{0s} + g_{\Lambda_b} n_{0s} n_b + g_{\Lambda_b} n_{\Lambda_b} n_{0s} + n_{\Lambda_b} n_{0s}, \tag{A.14}$$

and we can write the Eq. A.13 using the α as

$$s_{0,+} = \frac{n_{\Lambda_b} n_{0s} n_{tot}}{(n_0 + n_+)\alpha}. \tag{A.15}$$

For the B mesons (including B_{0s}), we get

$$c_{0s,m} = \frac{g_{0s} n_{\Lambda_b} n_{tot}}{\alpha}, \tag{A.16}$$

and finally for the b baryons (including Λ_B), we get

$$c_{\Lambda_b,b} = \frac{g_{\Lambda_b} n_{0s} n_{tot}}{\alpha}. \tag{A.17}$$

B. Details about the PDG reference values

All the PDG reference values used were from the Ref. [11]. Note that this PDG version is the 2020 version with 2021 update. In this appendix modified versions of Tables 5.1, 5.2, 5.3 and 5.4 are presented. These modified versions contain an additional column indicating the decay channel number Γ_x in the PDG.

Table B.1 contains the values of the admixture b decays. The PDG entry containing these values is named as "B±/B0/B0s/b-baryon ADMIXTURE". In PDG b and c denote hadrons. In this thesis a notation was used where b and c referred to quarks. Thus the decay modes $b \to c \to \mu/e$ and $b \to \bar{c} \to \mu/e$ are interchanged in Table B.1 compared to the PDG entry.

The decay channel number of the decay mode $b \to \tau \to \mu/e$ is marked in parenthesis as the PDG value shown in Table B.1, is calculated using BR($b \to \tau^+\nu_\tau$ + anything), corresponding to Γ_{25}, and the possible subsequent decays of τ via channels BR($\tau^- \to \mu^-\bar{\nu}_\mu\nu_\tau$), BR($\tau^- \to \mu^-\bar{\nu}_\mu\nu_\tau\gamma$), BR($\tau^- \to e^-\bar{\nu}_e\nu_\tau$) and BR($\tau^- \to e^-\bar{\nu}_e\nu_\tau\gamma$). The τ decays are from the PDG entry of τ and they correspond to decay channel numbers Γ_3, Γ_4, Γ_5 and Γ_6, respectively.

Table B.2 shows the exclusive semileptonic b hadron decay modes. The BRs are from the PDG entries of $B^\pm$, B^0, B_s^0 and Λ_b^0.

Table B.3 shows the inclusive BRs of b-to-c hadron decays. The BRs are from the PDG entries of $B^\pm$, B^0 and B_s^0. The modes with two possible final state D mesons, are calculated as the sum of the corresponding PDG modes. The decay channel number of modes $B^0 \to D^+$ and $B^0 \to D_s^-$ are marked with parenthesis, as only the upper limit for BR was given in the PDG. These modes were only used for the error calculation of the whole decay mode.

Table B.4 shows the b hadron production fractions. The BRs are from the PDG entry named as "B±/B0/B0s/b-baryon ADMIXTURE".

Table B.1: Semileptonic branching ratios of b hadron admixture.

Decay mode	PDG(%)	P8 avg.(%)	Γ_x
$b \to \mu/e$	10.69 ± 0.22	10.60 ± 0.01	7
$b \to \tau \to \mu/e$	0.45 ± 0.04	0.52 ± 0.01	(25)
$b \to c \to \mu/e$	8.02 ± 0.19	7.82 ± 0.01	27
$b \to \bar{c} \to \mu/e$	1.6 ± 0.5	1.6 ± 0.1	28

Table B.2: Semileptonic branching ratios for individual b hadron types.

Hadron	PDG(%)	P8 avg.(%)	Γ_x
$B^{\pm} \to \mu/e$	10.99 ± 0.28	11.29 ± 0.01	1
$B^0 \to \mu/e$	10.33 ± 0.28	10.44 ± 0.01	1
$B_s^0 \to \mu/e$	9.6 ± 0.8	9.3 ± 0.1	2
$\Lambda_b^0 \to \mu/e$	10.9 ± 2.2	7.7 ± 0.1	40

Table B.3: b-to-c hadron inclusive branching ratios.

Hadron	PDG(%)	P8 avg.(%)	Γ_x
$B^0 \to D^0/\overline{D}^0$	55.5 ± 3.2	46.03 ± 0.1	23/24
$B^0 \to D^+/D^-$	36.9 ± 5.1	50.8 ± 0.1	(25)/26
$B^0 \to D_s^+/D_s^-$	10.3 ± 3.3	12.9 ± 0.1	27/(28)
$B^+ \to D^0/\overline{D}^0$	87.6 ± 4.1	83.8 ± 0.1	38/39
$B^+ \to D^+/D^-$	12.4 ± 1.3	12.0 ± 0.1	40/41
$B^+ \to D_s^+/D_s^-$	9.0 ± 1.4	13.7 ± 0.1	42/43
$B_s^0 \to D_s^-$	93 ± 25	82 ± 1	1

Table B.4: b hadron production fractions.

Hadron	PDG(%)	P8 avg.(%)	Γ_x
B^+ (f_u)	40.8 ± 0.7	42.9 ± 0.1	1
B^0 (f_d)	40.8 ± 0.7	42.9 ± 0.1	2
B_s^0 (f_s)	10.0 ± 0.8	9.5 ± 0.1	3
b baryon (f_{baryon})	8.4 ± 1.1	4.7 ± 0.1	5

C. Supplementary plots to the methods and results sections

The supplementary plots are divided according to the different rescaling features they are related to.

C.1 b hadron production fractions

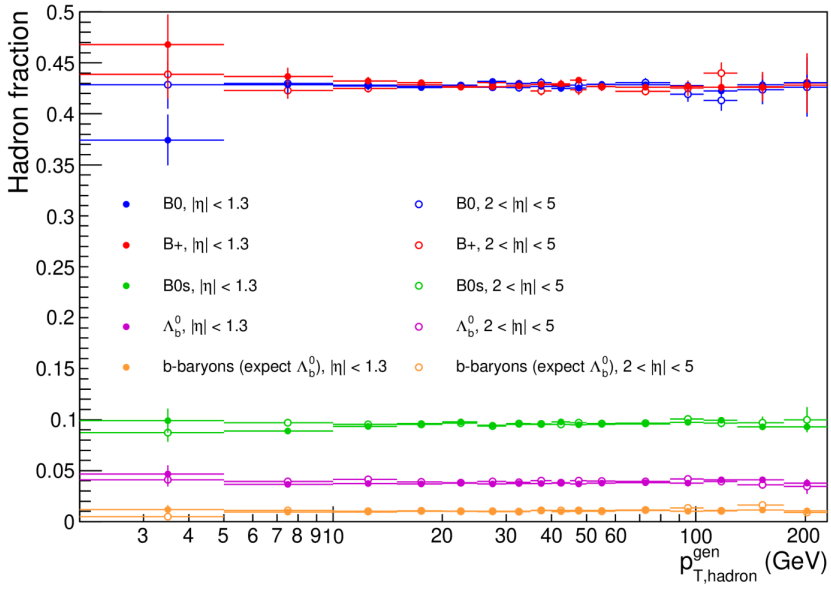

Figure C.1: Comparison of the b hadron production fractions in the η-regions used by the LHCb [66], $2 < |\eta| < 5$, and this thesis $1.3 < |\eta|$.

C.1.1 Original fit functions to production fractions

The original fit functions by the LHCb [66] are shown in Fig. C.2. The original fit function for $\frac{f_s}{f_u+f_d}$ is of the form

$$\frac{f_s}{f_u + f_d}(p_{\mathrm{T}}) = A(p_1 + p_2 \times (p_{\mathrm{T}} - \langle p_{\mathrm{T}} \rangle)), \tag{C.1}$$

where hadron p_{T} is referred to as p_{T}, $A = 1.0 \pm 0.043$, $p_1 = 0.119 \pm 0.001$, $p_2 = (-0.91 \pm 0.25) \cdot 10^{-3}$ GeV^{-1} and $\langle p_{\mathrm{T}} \rangle = 10.1$ GeV.

The fit function for $\frac{f_{\Lambda_b^0}}{f_u+f_d}$ is

$$\frac{f_{\Lambda_b^0}}{f_u + f_d} = A(p_1 + \exp(p_2 + p_3 \times p_{\mathrm{T}})), \tag{C.2}$$

where p_{T} again refers to the momentum of the hadron, $A = 1 \pm 0.061$, $p_1 = (7.93 \pm 1.41) \cdot 10^{-2}$, $p_2 = -1.022 \pm 0.047$ and $p_3 = -0.107 \pm 0.002$ GeV^{-1}.

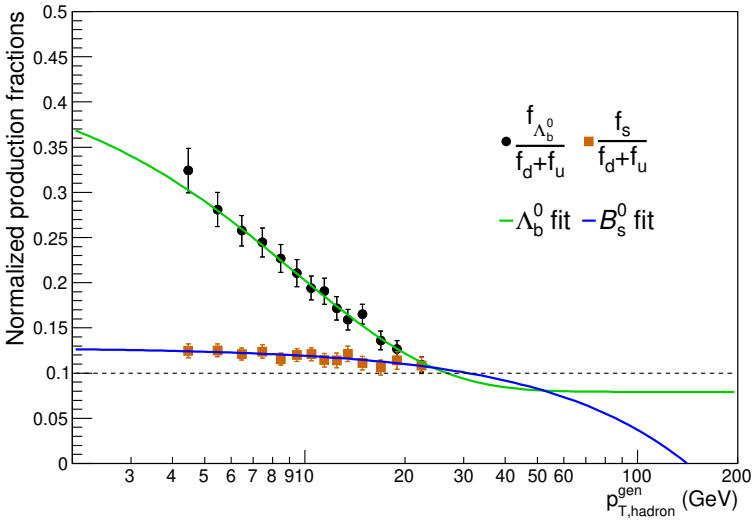

Figure C.2: Original fit functions used by the LHCb and the data which was used for the fitting [66].

C.2 b quark fragmentation

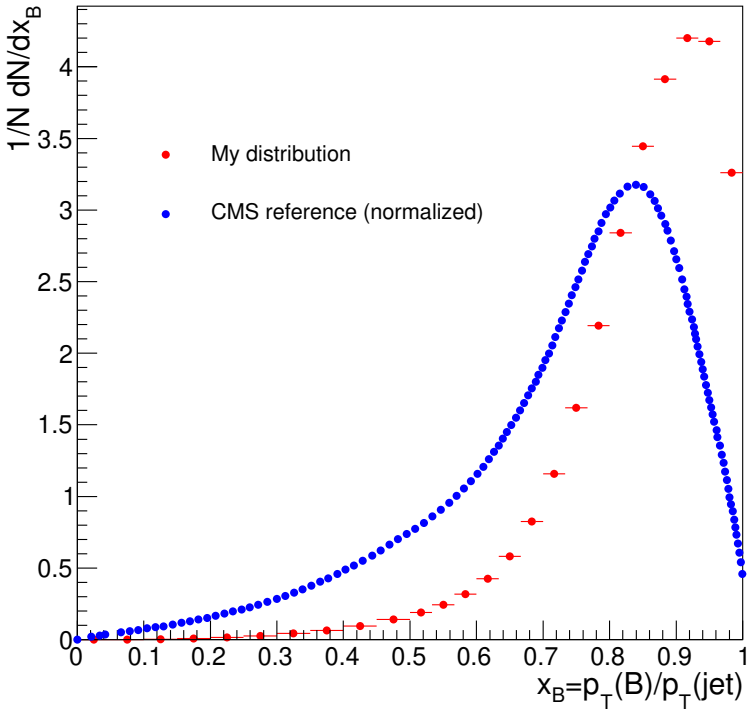

Figure C.3: Comparison of the x_B distribution produced for this thesis and the one presented in the CMS study [67].

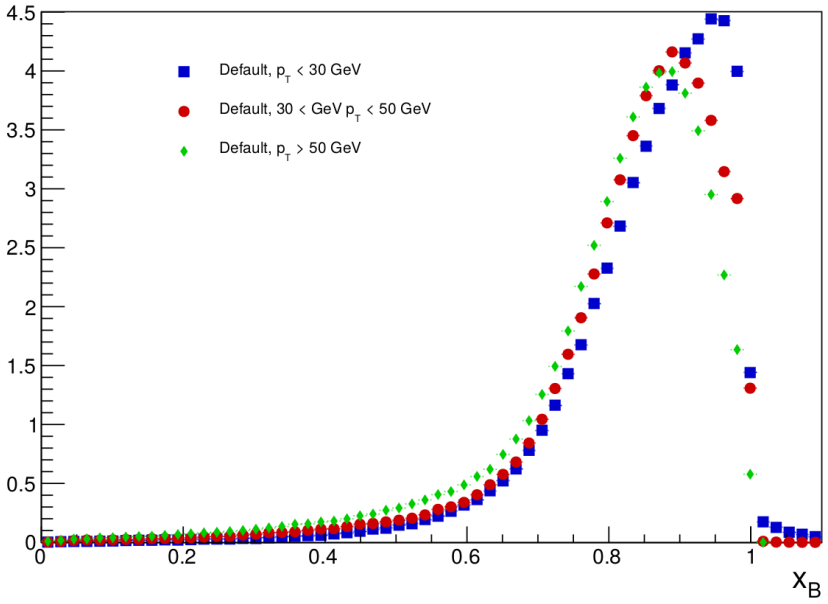

Figure C.4: The x_B distributions with different jet p_{T} cuts.

D. b-to-c hadron inclusive branching ratios with excluded scaling category

Figures D.1, D.2 and D.3 show the change in MC truth, MPF and p_{T}-balance responses as b-to-c hadron BR rescaling of either B^0 or B^+ was turned off. For comparison the normal b-to-c hadron BR rescaling, which was already presented in Chapter 6, is also included in the plots.

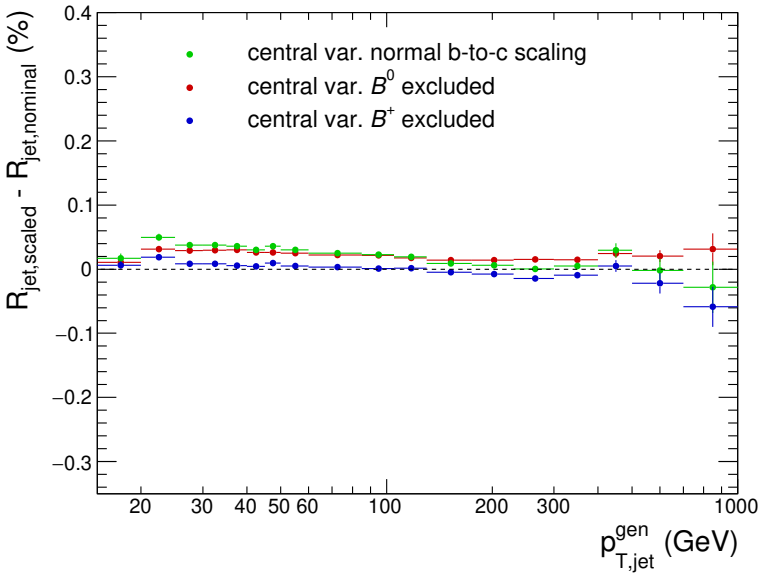

Figure D.1: Change in the MC truth response as b-to-c hadron BR rescaling of either B^0 or B^+ decays were excluded.

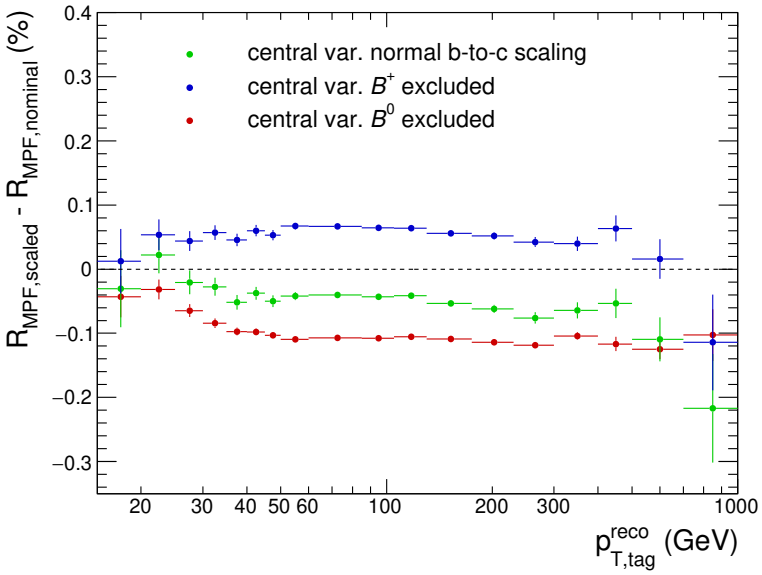

Figure D.2: Change in the MPF response as b-to-c hadron BR rescaling of either B^0 or B^+ decays were excluded.

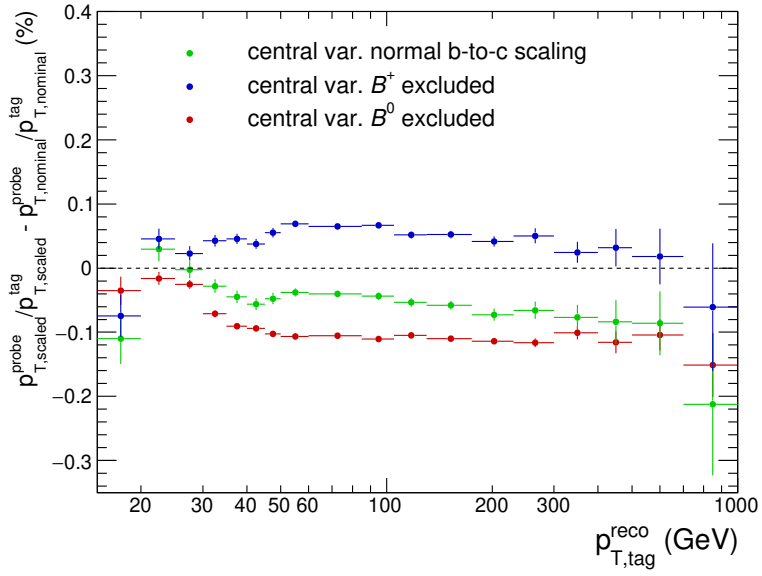

Figure D.3: Change in the p_{T}-balance response as b-to-c hadron BR rescaling of either B^0 or B^+ decays were excluded.

E. Plots with alternative total uncertainty estimate

This Appendix presents plots of Sections 6.1.5 and 6.2 with the alternative uncertainty estimation discussed in Section 5.4.5 using the exclusive b hadron semileptonic BR scaling.

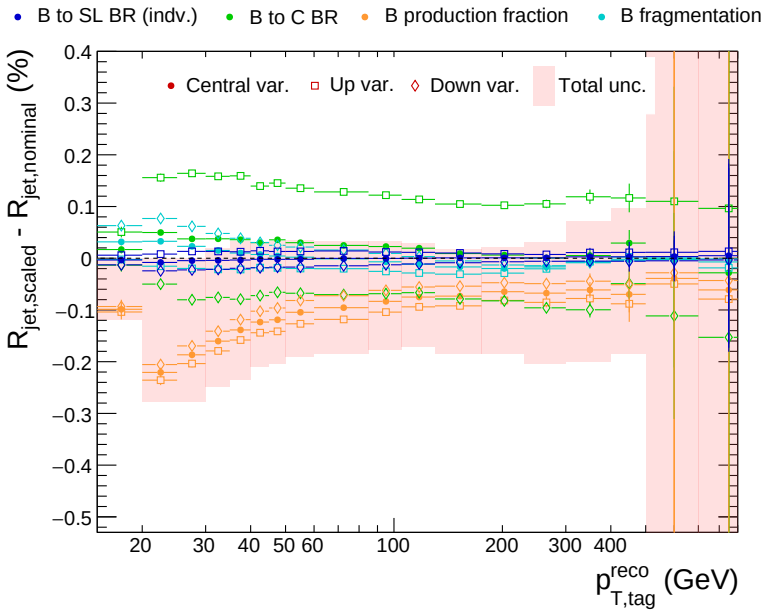

Figure E.1: Change in the b jet MC truth response due to the different scaling methods applied individually. The total uncertainty band on the background is calculated for the sum of all the scaling methods.

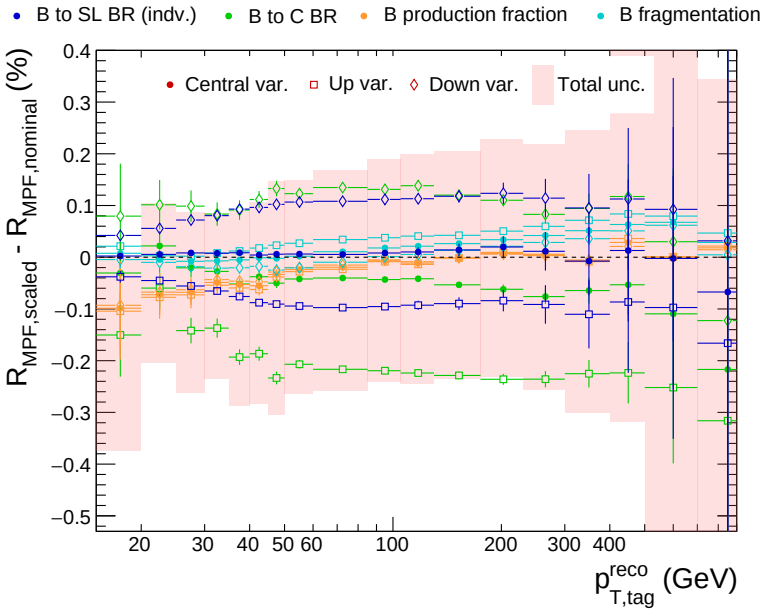

Figure E.2: Change in the b jet MPF-response due to the different scaling methods applied individually. The total uncertainty band on the background is calculated for the sum of all the scaling methods.

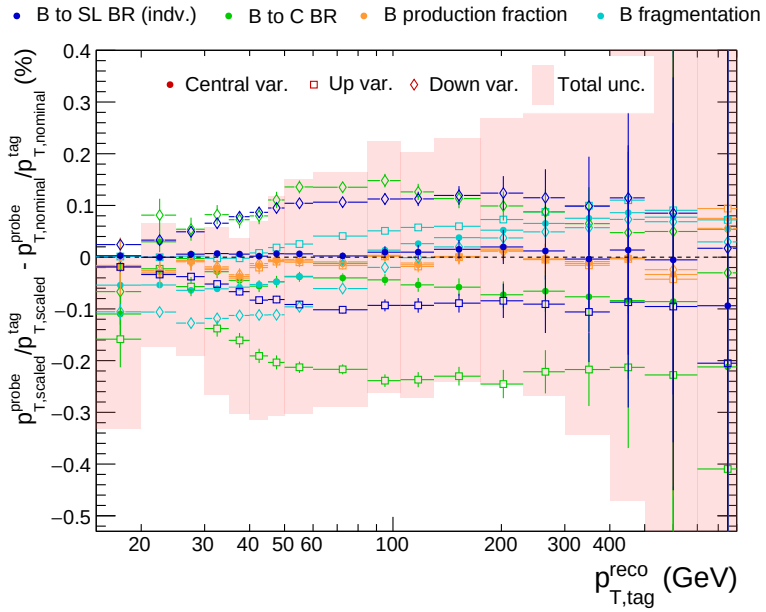

Figure E.3: Change in the b jet p_{T}-balance response due to the different scaling methods applied individually. The total uncertainty band on the background is calculated for the sum of all the scaling methods.

E.1 Total effect of the rescaling methods on the jet response

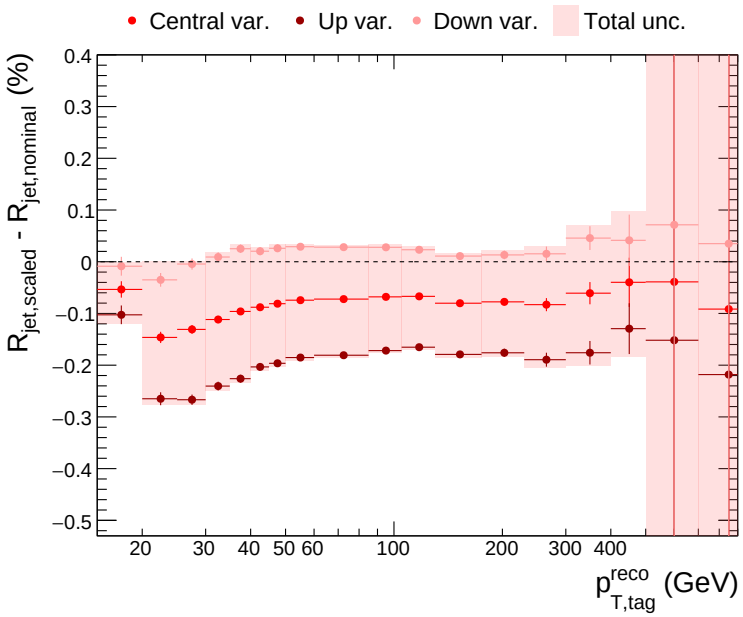

Figure E.4: Total shift in the MC truth response when all the scaling methods are applied.

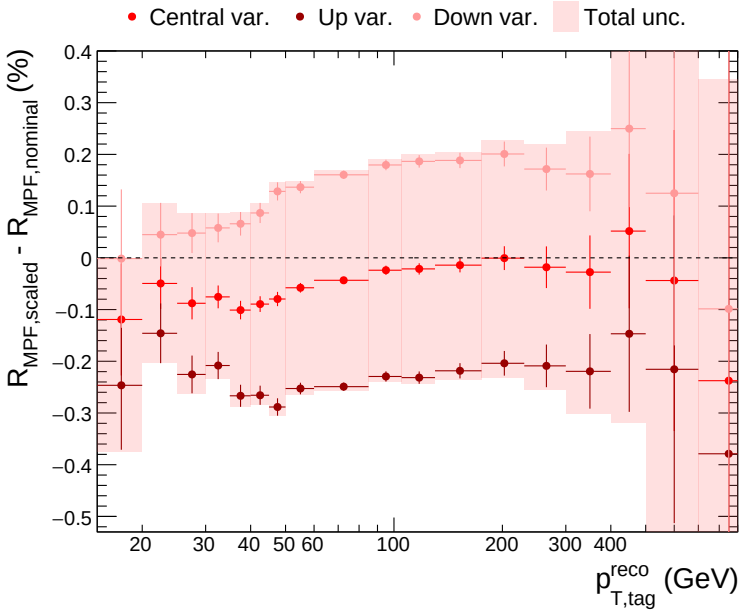

Figure E.5: Total shift in the MPF-response when all the scaling methods are applied.

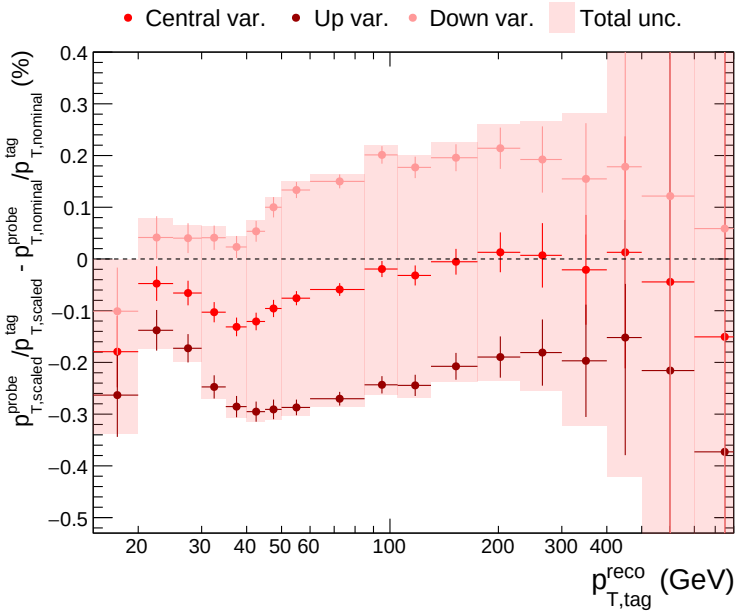

Figure E.6: Total shift in the p_{T}-balance response when all the scaling methods are applied.

F. Alternative plots with admixture b hadron semileptonic BR scaling

In the following plots, the admixture b hadron semileptonic BR scaling is used instead of the scaling by different b hadron types. The plots corresponds to the ones shown in Sections 6.1.5 and 6.2.

F.1 Summary plots of all the rescaling methods

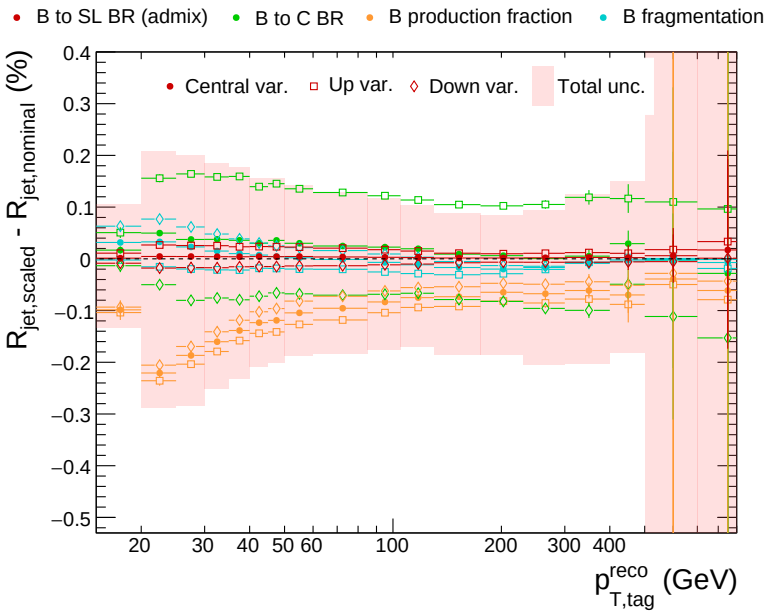

Figure F.1: Change in the b jet MC truth response due to the different scaling methods applied individually. The total uncertainty band on the background is calculated for the sum of all the scaling methods.

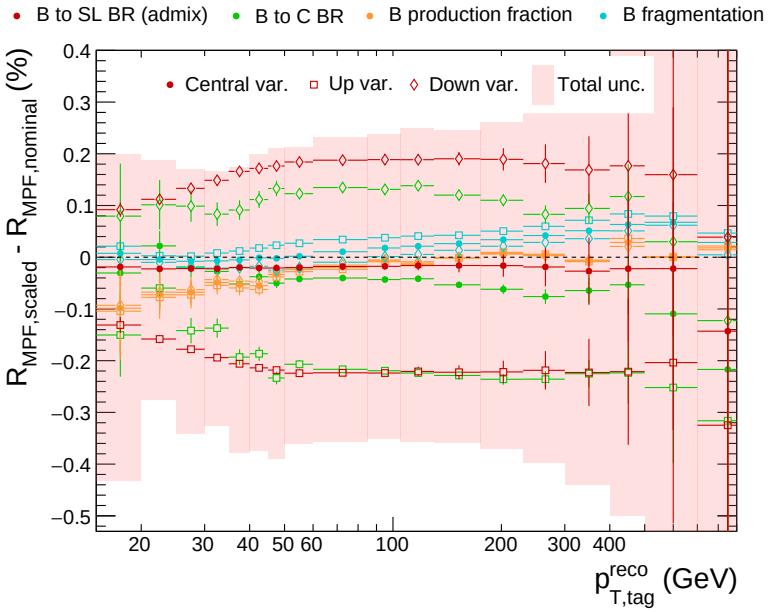

Figure F.2: Change in the b jet MPF-response due to the different scaling methods applied individually. The total uncertainty band on the background is calculated for the sum of all the scaling methods.

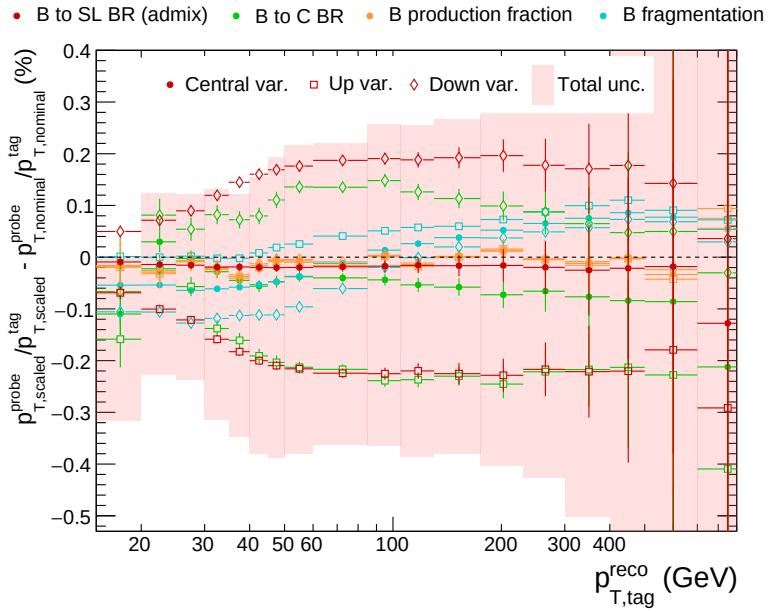

Figure F.3: Change in the b jet p_{T}-balance response due to the different scaling methods applied individually. The total uncertainty band on the background is calculated for the sum of all the scaling methods.

F.2 Total effect of the rescaling methods on the jet response

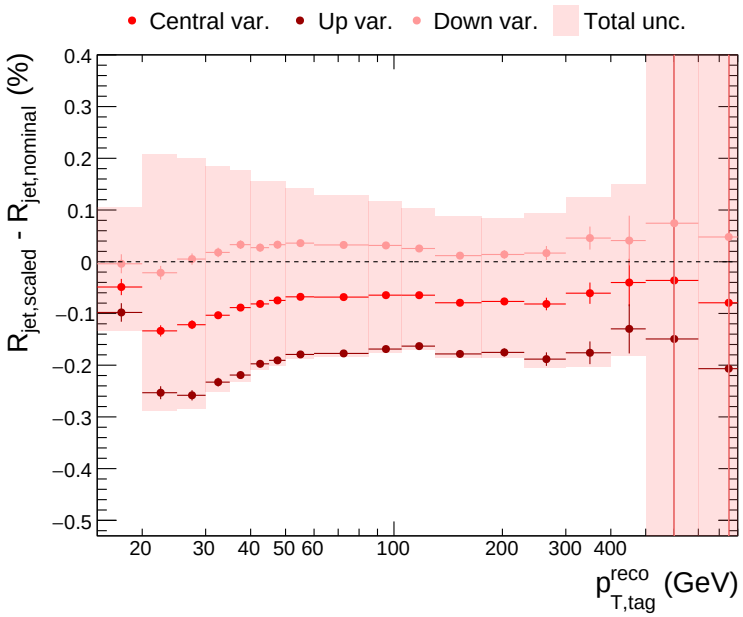

Figure F.4: Total shift in the MC truth response when all the scaling methods are applied.

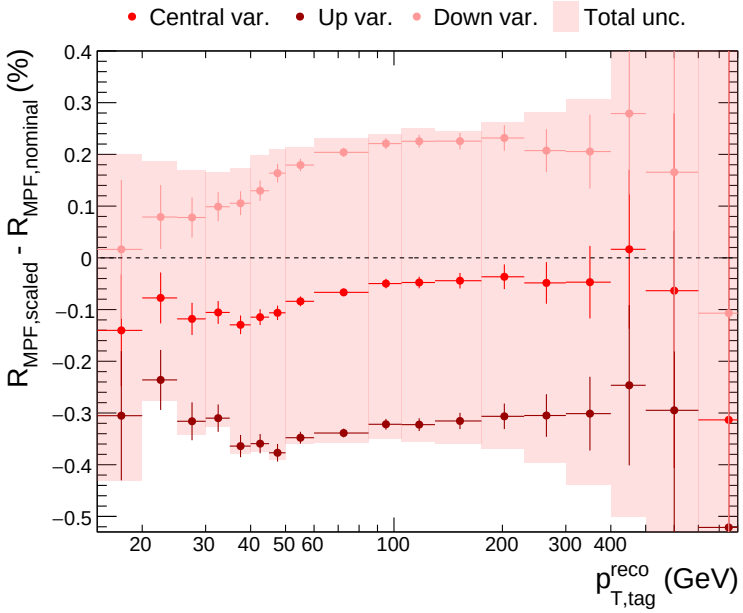

Figure F.5: Total shift in the MPF-response when all the scaling methods are applied.

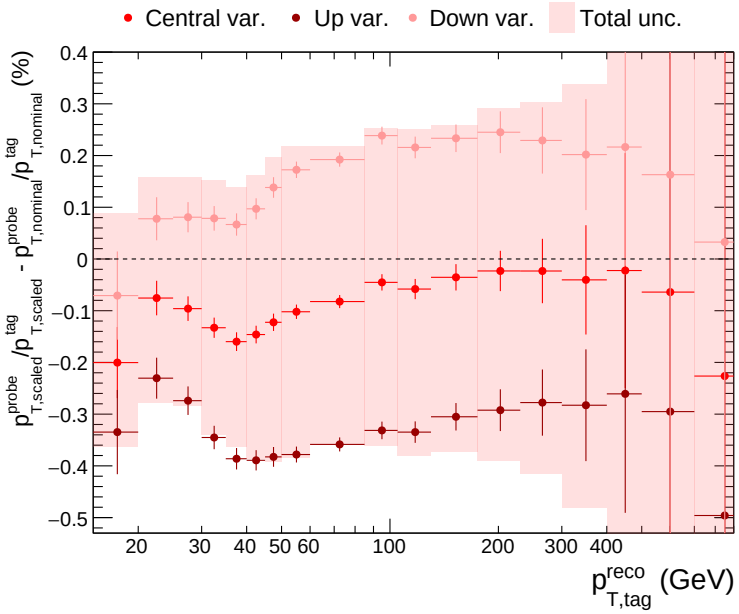

Figure F.6: Total shift in the p_{T}-balance response when all the scaling methods are applied.

G. Effects of individual scaling methods on the jet response with neutrinos added to MC jets

This chapter contains alternative versions of plots of Section 6.1. In following plots, the neutrinos are added to MC jets, and this will cause a shift in the MC truth response. The main motivation for this is to check whether the MC truth response will shift towards MPF- and p_{T}-balance responses. As can be seen from the plots of this section this is indeed the case. This means that the neutrinos make a major effect to the shapes of the MPF- and p_{T}-balance responses.

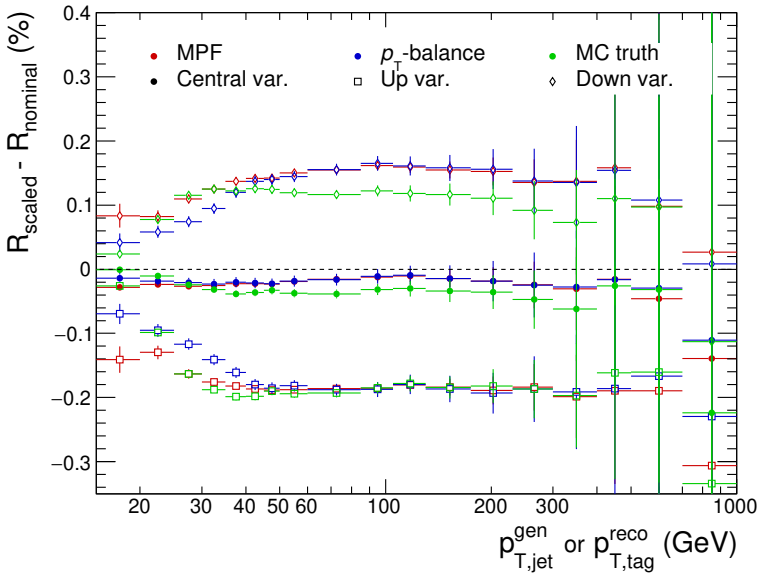

Figure G.1: Change in the b jet responses, when the semileptonic branching ratios of inclusive b hadron admixture decay channels were rescaled. Note that neutrinos are added to MC jets.

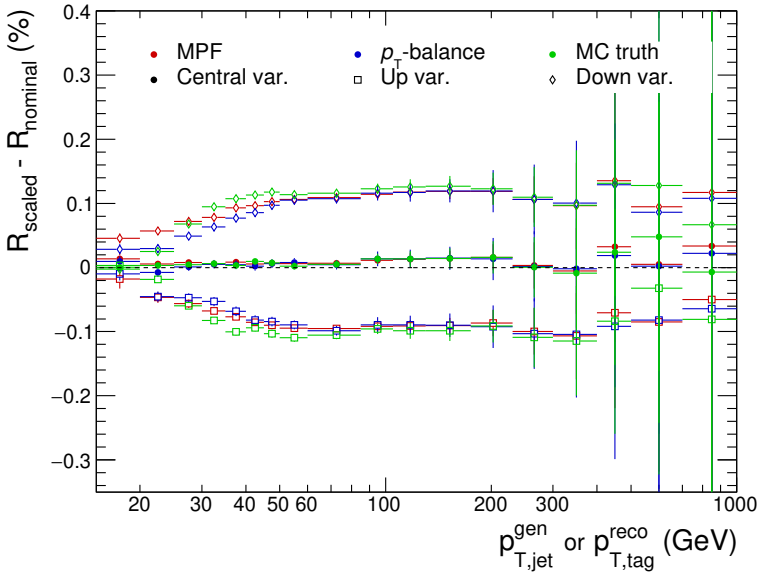

Figure G.2: Change in the b jet responses, when semileptonic branching ratios of individual b hadron types were rescaled. Note that neutrinos are added to MC jets.

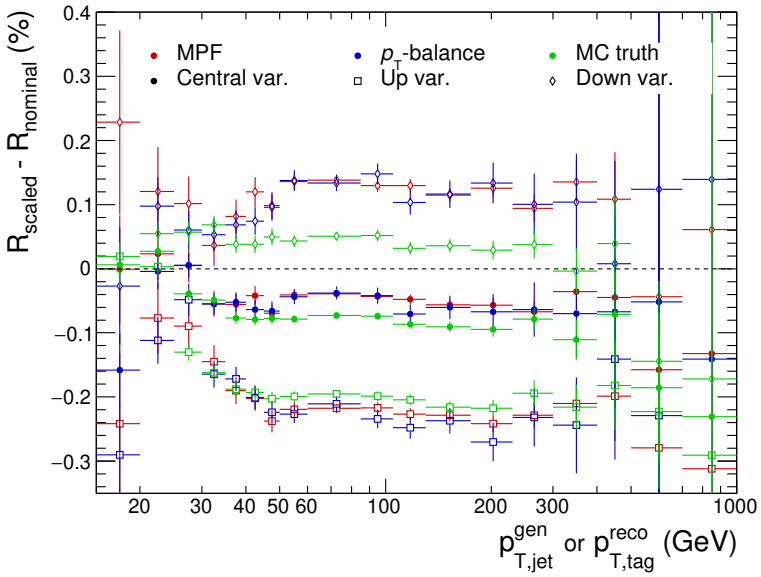

Figure G.3: Change in the b jet responses, when the b-to-c hadron inclusive branching ratios were rescaled. Note that neutrinos are added to MC jets.

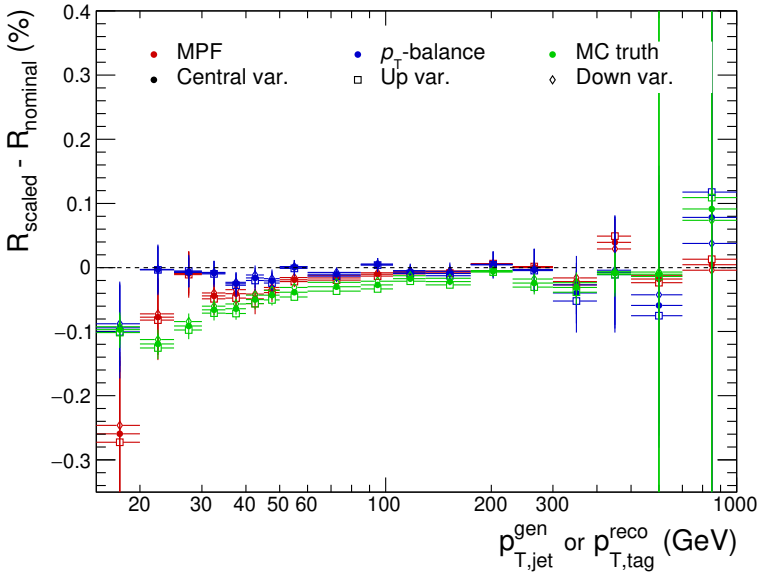

Figure G.4: Change in the b jet responses, when the production fractions of different b hadrons were rescaled. Note that neutrinos are added to MC jets.

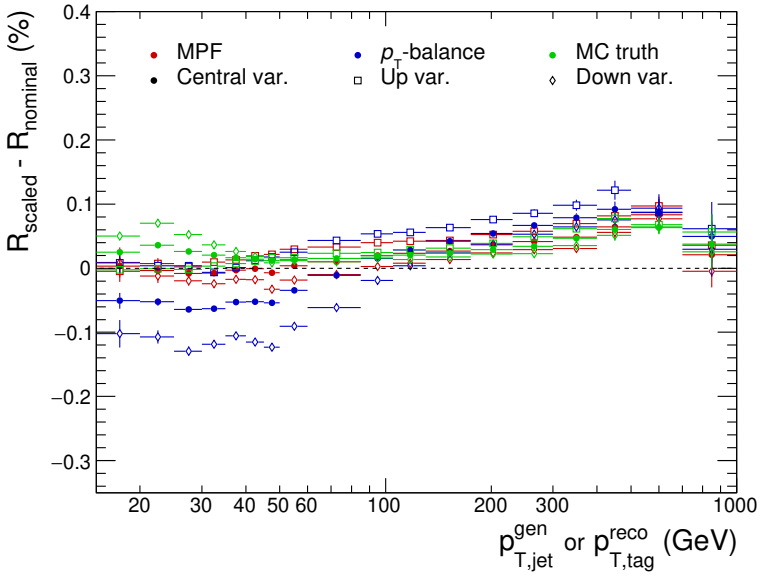

Figure G.5: Change in the b jet responses, when the b quark fragmentation function was rescaled to new parameterization. Note that neutrinos are added to MC jets.